

W

Waste Collection and Transport

JESS W. EVERETT

Civil and Environmental Engineering,
Rowan University, Glassboro, NJ, USA

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Source Activities
Collection
Transfer and Transport
Case Studies
Future Directions
Bibliography

Glossary

Source activities Actions that take place at the waste generation point that result in the proper presentation of the waste to the waste collection system.

Generation The act of creating a waste.

Separation Placing different waste materials in different containers.

Storage Keeping waste materials at the generation point until they can be collected.

Processing Changing the properties of a waste material, e.g., compacting cans.

Presentation Placing waste materials in the proper location for collection by the municipal waste collection program.

Participation rate The fraction of households that regularly present materials properly.

Residual waste Materials remaining as waste after reuse, recycling, or processing.

Set-out rate The fraction of households that present materials on collection day.

Curbside collection Collecting materials from each household, at the front curb or back alley.

Drop-off center A centrally located facility, with large containers for each type of material, from which material is collected.

Collection Obtaining materials from the curbside or drop-off centers and bringing that material to an unloading point.

Collection vehicle Vehicle designed to collect waste.

Transfer Moving wastes from a collection vehicle to a transport vehicle.

Transfer station Facility at which transfer is performed.

Material recovery facility Special type of transfer station, where recyclables are processed before transport.

Transport Moving wastes long distances for treatment, disposal, or recycling.

Transport vehicle Vehicle designed to move waste long distances.

Definition of the Subject and Its Importance

Concise definition of the subject and its importance with brief historical background.

Solid wastes must be collected from generation points and moved to a transfer station, or treatment, disposal, or recycling facility. Source activities are carried out at generation points. They include generation, processing, storage, and presentation to the collection system. Collection involves moving the waste from the presentation point to an unloading point, i.e., a transfer station, etc. Wastes are taken to a transfer station when the ultimate destination is far away. Environmental impact from collection and transport arise primarily from the operation of collection and transport vehicles.

Introduction

The functional elements of an integrated solid waste management system include generation, source activities,

collection, processing, transfer and transport, treatment, and disposal [1]. Generation is the creation of waste, when an item becomes unwanted. Source activities are things that must be done to the waste at the generation point to deliver it to the collection system, e.g., storage in a recycling or waste container that is placed at the curb on collection day. Collection involves obtaining the waste from the presentation point and bringing it to the unloading point of the collection vehicle. Processing is anything that must be done before transfer, transport, or disposal. Transfer moves the waste from the collection vehicle to a transport vehicle, perhaps via a processing facility. Transport involves moving waste significant distances from a transfer location to processing, treatment, or disposal facilities. Treatment facilities reduce the volume, mass, or toxic nature of a waste. Treatment facilities include composting facilities and incinerators with energy recovery. Disposal facilities store wastes, e.g., landfills. All landfills combine aspects of treatment and disposal, with bioreactor landfills tending more toward treatment than traditional landfills.

The purpose of this entry is to describe the collection and transport of waste, with emphasis on associated environmental impacts. The functional elements most related are source activities, collection, transfer, and transport.

Source Activities

Source activities take place at the point of waste generation, i.e., residences, commercial, industrial and institutional establishments, and outdoors (picnics, camping, travel, etc.). Source activities must be carried out properly if waste is to be properly delivered to the solid waste collection system. Source activities include separation, storage, processing, and presentation. For the collection of recyclables and compostables, the participation rate is also very important.

The environmental impacts of source activities are relatively small and have not yet been subjected to analysis or estimation. The discussion here is focused on their description.

Separation

Separation requirements can vary depending on the solid waste management system. If no recycling or

Waste Collection and Transport. Table 1 Common residential separation schemes

Scheme	Description
Wet/Dry/Residual	The wet stream is composted while the dry stream is incinerated. Residual wastes are landfilled. Recyclables are often separated from the residual wastes via a drop-off center program.
Single-stream recycling	All recyclables (paper, bottles, cans) are separated into a single stream, which is further separated at a central processing facility. Residual wastes are also separated. Compostable wastes may be separated from residuals.
Two-stream recycling	Paper and cardboard are separated into one stream, while bottles and cans are separated into another. The bottle and can stream typically consisted of plastic, glass, aluminum, and steel. Residual wastes are also separated. Compostable wastes may be separated from residuals.

composting is included in the system, no separation is required; all wastes can be placed in a single container. A single container can also be used if a centralized facility is used to separate recyclable or compostable materials from residual waste.

In many solid waste management systems, waste must be separated into two or more fractions. Common residential separation schemes are described in Table 1.

The separations employed at industrial and institutional establishments will depend on the waste characteristics of a particular facility. Very specific separations may be justified, depending on the predominance of a particular type of waste, e.g., grocery stores may separate corrugated boxes, offices may separate white paper, and cafeterias may separate food waste.

Storage

A two-step storage process is often followed in residences. Small containers are kept in the kitchen (under the sink or in a cabinet) or in a laundry or utility room. When these small containers become full, they are emptied into larger containers kept

outside the living area. At single-family dwellings, the containers are often kept next to the house, along an alley if present, or next to or in a garage or outbuilding. In the case of compostables, materials may be stored for a brief time in the home and then moved to an outside storage container or a home composting unit.

At multistory apartment complexes garbage chutes may be available in the hallways. In some cases, these chutes can accept multiple fractions. A switch is used to identify the fraction about to be placed in the chute and a carousel rotates the proper container under the chute. Alternatively, convenience areas may be located outside the apartment building. Convenience areas may contain one container (for the entire waste stream) or multiple containers for recyclables, compostables, and residual waste. An apartment complex convenience center in Börlange, Sweden contains separate containers for batteries, compostables (wet), combustibles (dry), residual waste, colored glass, and clear glass.

Storage at industrial or institutional establishments may also be two-stage. For example, at an office facility, small containers may be located in offices and/or public areas. These containers will be periodically emptied into larger containers in utility areas or outside.

A relatively small number of municipal solid waste (MSW) systems are pneumatic. In this case, containers may be located inside or near buildings. These containers are periodically emptied by suction and the materials moved through pneumatic tubes to larger containers for longer-term storage.

Processing

Home composting is a relatively common process employed at residences. Composting is controlled aerobic or anaerobic biodegradation in which microbes consume the more biodegradable portions of a waste, though worms can also be used. Backyard composters can range from a simple enclosure of wire fence to a container that is rotated to mix and aerate. Worm composters consisting of a tower of trays can be located inside the home. Fresh waste is deposited in a top tray, causing worms to move out of composted material in a lower tray, leaving it worm free and ready to be removed. Yard and food waste are the most common materials composted at home. Maximum savings to communities are realized from home composting, as

a portion of the waste stream is treated at zero cost to the municipality. Even more convenient is the “let it lie” process, in which leaves or cut grass are allowed to remain in the yard and degrade naturally.

Most processing at the source is carried out to make it easier to store waste. Some items must be cleaned before placement in a recycling container, such as a mayonnaise jar. Other items may be crushed by hand to increase density and store more items before a container must be emptied. In some houses an appliance is used to compact residual waste, to minimize the frequency with which it must be taken to the outside storage container.

Processing may be required at industrial or institutional establishments, e.g., at a grocery store, corrugated boxes may be compacted to optimize storage.

Presentation

The last source activity is “presenting” the waste to the collection system. If curbside collection is employed, a collection vehicle passes each single-family residence. In some cases, the collection crew may walk onto the property to retrieve container(s). The crew may even return the container to its storage location. In this case, the resident does not need to do anything other than fill the container. More commonly, residents move their containers to the curb or alley on collection day and/or return them to the storage location after collection.

Where a large centralized container is located near an apartment complex or industrial or institutional establishment, no special presentation is required. Simply storing the waste puts it in the proper place for presentation to the municipal solid waste system.

In communities that employ drop-off centers, large centralized containers are located at a relatively small number of centrally located public areas, e.g., along streets or at grocery stores, government buildings, transfer stations, or landfills. Residents are responsible for bringing materials to these drop-off centers, where they are stored temporarily before collection. For the residents of the Börlange apartment complex mentioned above, additional materials were presented to the MSW system at a nearby grocery store. Containers included in-store storage for deposit/refund beverage containers (glass, plastic, and aluminum), and storage containers in the parking lot for packaging (paper,

plastic, and metal). While drop-off centers are less expensive to operate for the municipality, locations should be picked so that users can walk or, if driving is required, use combined trips. Otherwise, the economic and environmental costs incurred by the generators may exceed any municipal savings.

According to Belton et al. [2], drop-off centers should be within 500 m if an individual is walking or 0.5 to 2.5 km if the person is driving.

Valeo et al. [3] used geographical information systems (GIS) to locate drop-off centers for recyclables in a community of 22,000 in Southern Ontario, Canada. Location-allocation modeling was used. Shopping centers, municipal parking lots, and roadside sites were considered. Five-hundred potential sites were identified. If people were assumed to walk to drop-off centers, 35 centers were needed to adequately serve 99% of the population.

Bacha et al. [4] developed a model predicting the collected amount of waste paper based on 649 Austrian municipalities using drop-off centers. Municipalities with a higher number of overnight hotel stays per person and year had higher quantities of collected waste paper, attributed to the increased waste potential of municipalities in tourist regions. The index of purchasing power per capita (the sum of net income per year) showed a strong effect on the amount of waste paper collected. Municipalities with a higher percentage of households in which one member works in the home for no pay showed a significantly lower amount of waste paper per person. Higher amounts of waste paper were collected in municipalities with a higher percentage of employees in businesses and industry as well as in the service sector. A higher number of agricultural firms were associated with a smaller amount of collected paper. The density of collection sites positively influenced the amount of collected waste paper, probably due to increased convenience.

Surveys can be used to improve the operation of drop-off centers. Williams and Taylor [5] administered two surveys in order to better understand how to maximize the use of 26 drop-off centers in Lancashire, UK. A telephone survey was carried out among the drop-off center attendants to identify the effects of recent site improvements and obtain their opinions regarding overall customer satisfaction. An on-site questionnaire survey was conducted among customers

at five drop-off centers, over three seasons, on weekends and weekdays. Customers were asked their opinion of the drop-off centers and were asked to identify methods that would maximize recycling. Site attendants stated that they actively encouraged customers to segregate waste more efficiently, and that a bonus scheme had provided incentive regarding this effort. Site attendants had a key role in effective use of the drop-off centers. Lancashire drop-off center customers were mostly male, used private cars, and showed high levels of customer satisfaction. Practical changes to encourage more recycling at drop-off centers included longer opening hours and the provision of boxes to enable pre-separation of waste.

Participation

In residential areas, the participation rate in collection programs is the fraction of residences that present all or most of their waste materials (residual, recyclable, or compostable) to the collection system. Typically, 100% of residences present residual waste, because municipal collection is the most convenient way to get rid of a material that otherwise will cause odors and attract flies. Smaller fractions of residences participate in municipal recyclables or compostables collection programs, because it is more convenient to leave them in the residual fraction.

The set-out rate is the fraction of homes that set out materials on a given collection day. Set-out rates are lower than participation rates because some participating households do not set out materials on every collection day. Set-out rates for residual waste programs are also close to 100%, while set-out rates for recycling programs vary greatly. According to Canadian researchers, set-out rates for recycling programs have been reported to range from less than 20% to more than 70% [6].

Given that it is typically more convenient to leave recyclables and compostables with residual waste, it is useful to investigate reasons why significant numbers of people participate in recyclables and compostables collection programs. Recycling and composting programs have been described as individual-level environmental collective actions (IECAs). According to Everett [7], IECAs involve the widespread adoption of particular individual behaviors to produce environmental collective

goods. Collective goods are defined as any good “such that, if any person X_i in a group $X_1, X_2, \dots, X_i, \dots, X_n$ consumes it, it cannot be feasibly withheld from others in the group” [8]. Some collective goods have what is called “jointness of supply.” For collective goods with jointness of supply, the consumption of the collective good by one member of the group in no way reduces the ability of other members to enjoy it. The classic example is the lighthouse. Once one is built, any ship passing near it benefits (a collection good), and the benefit of one ship does not reduce the benefit for other ships (jointness of supply).

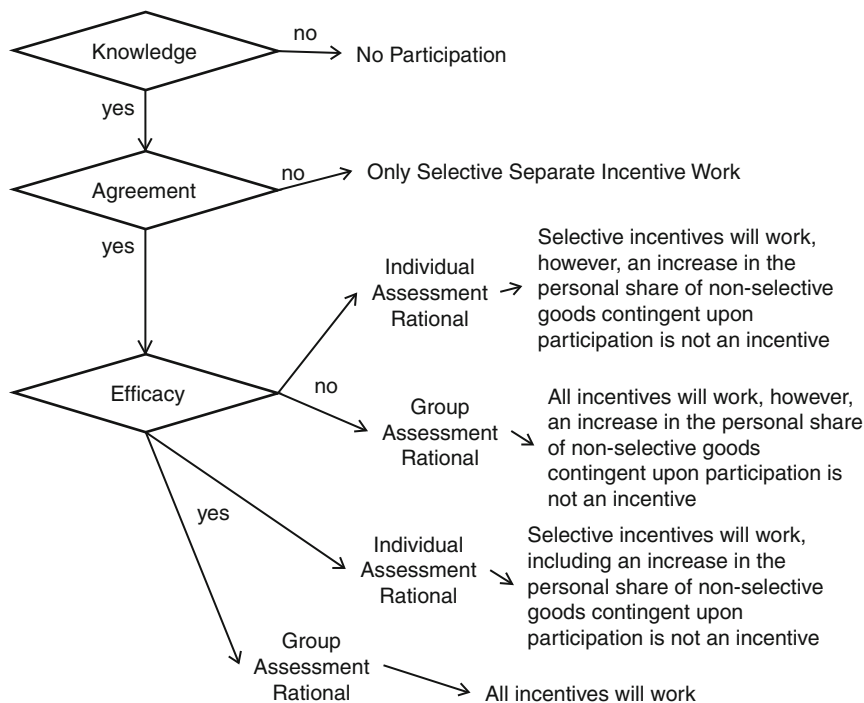
Collection programs are IECAs because they require individuals to adopt behaviors concerning municipal solid waste management in order to produce environmental collective goods such as cleaner air, water, and soil. Of course, not all of the goods provided by collection programs are collective in nature, e.g., lower MSW management bills.

Figure 1 can be used to understand the factors leading to participation in IECAs in general and

recycling and composting programs in particular [7]. The Figure provides a flowchart that describes the various scenarios under which various assessment rationales and incentives are effective. In the discussion that follows, the Figure is interpreted for recycling programs; however, the discussion can easily be applied to composting programs.

The first condition is knowledge of the recycling program. People who do not know that a recycling collection program exists will not participate. Furthermore, while it is unlikely that a curbside collection program can remain unknown for a long time, residents that do not learn what or how to recycle may not participate. With regard to drop-off programs, residents may remain ignorant of the existence of a program, especially when drop-off centers are located in places rarely visited for purposes other than recycling.

Once a resident has knowledge of the recycling program, participation requires sufficient incentive. Incentives can be described by two dimensions,



Waste Collection and Transport. Figure 1

Conditions to participate in recycling programs [7]

separation and selectivity. Separate incentives do not depend on ideological agreement with the goals of the recycling program, e.g., monetary incentives. Not-separate incentives are related to the goals of the recycling program, e.g., saving natural resources or reducing waste management costs. Selective incentives are contingent on participation, e.g., awards or avoiding punishment. Nonselective incentives are the collective goods provided by the IECA. Anyone in the group receives them, whether or not they participate.

There are four combinations of these two dimensions: selective-separate (e.g., monetary rewards); selective-not-separate (e.g., awards); nonselective-not-separate (collective goods directly related to the goals); and nonselective-separate (collective goods supplied by the recycling program, but not related to the goals). The last combination is rarely encountered.

The effectiveness of each type of incentive depends on ideological agreement, assessment rationale, and belief in efficacy. Individuals in ideological agreement with a recycling program believe that its goals are valid. Only selective-separate incentives work for individuals not in ideological agreement.

For those that agree with the goals of an IECA, their perception of individual efficacy and their assessment rationale jointly determine what incentives will have influence. With many collective actions "...no single individual's contribution makes a perceptible difference to the group as a whole, or the burden or benefit of any single member of the group," i.e., individual efficacy is negligible [8]. For IECAs, even though an individual's contribution to the overall collective good may be small, it exists and its production is directly linked to participation. As each individual's ability to produce the collective good, though small, is approximately the same, each individual has about the same effectiveness. Some individuals focus on the direct relationship between their participation and production of the collective good and end up believing in their own efficacy. Others focus on the small amount of collective good their participation produces, and believe that they are not efficacious.

Two assessment rationales can be employed, individual and group. An individual assessment rationale is based only on personal gain from participation. Collective goods, the receipt of which does not depend on participation, enter into individual assessments only if

the *increase* in an individual's share of a collective good consequent to participation is greater than the cost of participation. Such an individual, by Olson's terminology [8], is "privileged." It is difficult to imagine such a person with respect to collection programs. If greater profit – personal benefit minus personal cost – is gained from not participating, nonparticipation will be the choice of action of individuals operating under an individual assessment rationale. This is the logic behind "free riders."

Individuals following a group assessment rationale consider group benefits. Some employing the group assessment rationale will participate any time the group benefits outweigh their personal cost. Since the small improvement in environmental quality from recycling a single item is shared by all living beings on the planet, it is very likely that the group benefit will outweigh the individual cost. Others employing the group assessment rationale may not be so altruistic; they will participate as long as their benefits outweigh their personal costs. They differ from people employing the individual assessment rationale because they may participate when nonparticipation would increase their personal net benefit. Figure 1 identifies the types of incentives that influence people falling into the four combinations of efficacy and assessment rationale.

At least five major factors are related to participation in collection programs: market incentives, coercive incentive, convenience strategies, promotional efforts, and demographic and attitudinal variables [7]. Market incentives are selective and separate. They consist of direct payments, deposit programs, and pay-as-you-throw (PAYT) schemes. In communities using PAYT, residents pay different rates for different volumes or mass of waste collection. Pay-as-you-throw programs are an effective method for encouraging source reduction and recycling. The US EPA [9] estimated that each person participating in a PAYT program reduced their annual greenhouse gas (GHG) emissions by an average of 0.085 MTCE. An MTCE (metric ton carbon equivalent) is an amount of greenhouse gases (e.g., CO₂, Methane, water vapor, or ozone) with the greenhouse potential of an amount of CO₂ containing 1 metric ton of carbon.

Coercive incentives use the force of law to produce desired behaviors. A nationwide study in the USA found that mandatory programs collecting newspaper,

glass, and aluminum collected almost 60% more materials than voluntary programs collecting the same materials [10]. Though numerous programs have rules requiring participation, and even punishments specified in ordinances and codes, enforcement is rare.

Though separating wastes into additional fractions for recycling or composting can never be more convenient than placing everything in one container, convenience strategies can be employed to reduce that additional cost as much as possible. Convenience strategies can employ delivery methods, separation requirements, container provision, collection frequency, and program reliability [10, 11].

Promotion efforts give people knowledge of a program, describing where, when, what, and how to recycle. They can also affect ideological agreement and efficacy, e.g., by advertising recycling benefits and describing the positive outcomes of an individual's recycling behavior. Promotions can also change people's perceptions of all types of incentives, e.g., lowering personal cost perceptions by describing the convenience of a particular program. Promotions should be tailored for each program, highlighting its particular benefits and conveniences, and minimizing its costs.

The explanatory power of demographic and attitudinal variables can be related to recycling activity [7]. Demographic variables such as higher education, higher income, neighborhood stability and networks, type of building, and age group have all been associated with recycling behavior. So have attitudinal variables such as a sense of being in control of one's life, a sense of personal responsibility for environmental problems, and frugality.

Collection

Collection is needed to move wastes from widely distributed generation points to a transfer station or treatment/disposal facility. In the case of residual wastes, collection frequency has an upper limit based on the breeding cycle of pests, such as maggots and flies. For example, it takes 9 to 11 days for flies to multiply in residual waste at typical temperatures [1], hence residual wastes are generally collected at least once a week. While recyclables can typically be stored for at least 2 weeks, the once per week collection frequency is

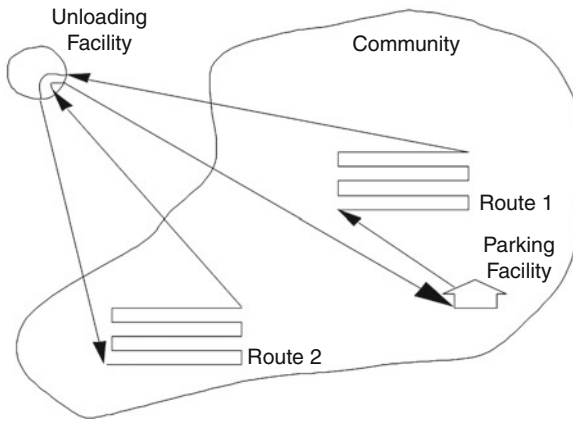
often used, as it is easier to remember to collect every Wednesday, e.g., rather than every other Wednesday.

Collection programs can be privately or publicly operated. García-Sánchez [12] examined street cleaning and waste collection services in Spanish municipalities to identify factors influencing service efficiency. Information from 38 of 113 towns with populations over 50,000 was obtained. Measures of success included tonnage collected, number of collection points, collection point density, and length of streets washed. They found no significant difference in the efficiency of street cleaning and waste collection services between towns with private versus public programs.

Collection containers can be divided into two major types, hauled and stationary. Full hauled containers are carried by the collection vehicle to the unloading point. Once emptied, they are returned to their next collection point. A common use for hauled containers is at construction and demolition sites. Stationary containers are permanently located at a particular site. When full, or at regular intervals, the stationary container is emptied into the collection vehicle and immediately returned to its location.

Stationary containers range from relatively small containers used in curbside collection programs to specialty containers and dumpsters. In curbside collection programs, vehicles drive by each setting out household or business and collect materials. Curbside programs most commonly collect from neighborhoods of one- and two-family residences or districts of small businesses. Typical curbside containers in the USA range from 32 to 96 gallons (121–363 L). Figure 2 is used to demonstrate the typical activities of a collection vehicle collecting two routes per workday. The vehicle is stored in a parking facility. At the start of the workday, the truck is driven to the first route and collection begins. When the route is finished, the truck drives to the unloading facility, unloads, and travels to the second collection route. Once the second route is finished, the truck again travels to the unloading facility, unloads and returns to the parking facility. Unloading facilities can be processing centers, transfer stations, landfills, or incinerators.

Dumpsters and specialty containers are located at large generation sites (e.g., apartment complexes and businesses) or at drop-off centers. In communities that rely on drop-off centers, householders must take



Waste Collection and Transport. Figure 2
Collection truck activities (Residual waste, recyclables, or compostables)

recyclables or compostables, and, in some cases, residual wastes to containers located at public access points, e.g., along streets in urban areas, shopping center parking lots, municipal facilities, transfer stations, or landfills. Specialty containers are often used at drop-off centers, e.g., containers with openings designed to accommodate different recyclable items. Collection vehicles periodically empty the containers.

The collection of residual or compostable waste is most often done in compaction vehicles. By compacting waste during the collection process, more waste can be collected before the vehicle must be unloaded. Compaction can be used with recyclables, but care must be exercised to avoid breaking glass.

Vehicle loading can be manual, semiautomatic, or automatic. Manual loading requires the collection crew to push or pull containers to the collection vehicle and lift them to empty. Semiautomatic still requires a crew member to push or pull the container to the collection vehicle, but a pneumatic lifting system is used to empty the container into the truck. In automatic collection, the driver of the truck maneuvers close to a container, “grabs” it with a crane or “forklift” type device, and automatically empties it. Automatic systems use the crane with curbside collection and the “forklift” type device with dumpsters. Single compartment trucks are the norm for residual waste collection. The collection of recyclables can be done in single- or multi-compartment trucks.

Manual and semiautomatic collection can result in health issues for crew members. Rogers et al. [13] administered a survey to municipal solid waste (MSW) collectors and their supervisors/safety officers in both public and private companies to evaluate occupational risks to Florida MSW workers. Data were obtained for 251 workers. Results of the survey indicated high rates of lacerations, contusions, strain/sprains, and illness. Seventy-five percent of the collectors reported having been injured in the previous 12 months.

According to Medina [14], many cities in developing countries spend 20–50% of municipal revenues on MSW while collecting only a fraction of MSW generated, in many cases less than 50%. In Mexico, informal private collectors often provide collection in low-income areas, where no official collection is provided. While informal refuse collection is illegal in Mexico, response varies from city to city, including repression, neglect, collusion, and even encouragement. Medina argues that informal refuse collection can be a valid part of solid waste management where a regulatory and policy framework is in place that encourages self-help, private investment, and entrepreneurship. Many cities in developing countries are simply unable to provide complete collection coverage, especially in low-income areas. Informal refuse collection can fill the gap, providing an important service and improving the quality of life.

Factors Influencing Collection Efficiency

WRAP [15], a nonprofit based in the UK, completed an analysis of three types of curbside collection systems for recyclables: curbside sort; single stream comingled; and two stream partially comingled. In curbside sort systems, materials are sorted by type at the curbside and collected in different compartments of a multi-compartment vehicle. In single stream comingled systems, materials are placed at the curb in one bin and collected in a single compartment vehicle, with the sorting of the materials occurring at a Materials Recovery Facility (MRF). In two stream partially comingled systems residents separate materials into two categories, usually fibers (paper and cardboard) and containers (glass, cans, and plastic bottles), which are collected in a two compartment vehicle. Each comingled stream is handled separately at the MRF.

The following general observations were made after investigating programs in the UK:

- Curbside sort schemes had lower costs than single stream comingled schemes, when the income from material sales was considered.
- The net costs of curbside sort schemes are strongly affected by income from the sale of materials.
- The net costs of comingled schemes are strongly affected by MRF fees.
- Two stream comingled collections schemes keeping paper separate from containers have similar net costs to curbside sort schemes.
- There is little variation in material yields between the three main curbside schemes.
- Programs that collect glass and/or have alternate weekly collection of refuse exhibit the greatest diversion rates.
- Recycling collections are maximized when more or larger containers are provided and/or recyclable materials are collected weekly.
- The best scheme does not appear to be impacted by the “urban” or “rural” nature of the area served.

Wilson and Batz [6] developed a derived probability model to estimate vehicle and labor requirements for municipal solid waste collection systems. The set-out distribution was modeled using the binomial distribution, leading to the application of the geometric distribution to stop-to-stop distance. The total loading time and delay time were modeled using the normal distribution. The number of households a truck passes before filling was modeled using the negative binomial distribution. The model was used to explore three scenarios: route size when the collection vehicle is constrained by time; route time when the vehicle is constrained by capacity; and the effect of vehicle capacity on costs [16]. Parameters included route size (the number of houses served before the truck became full or runs out of time), set-out rate (the fraction of houses setting out materials on a given day), average loading time (time to load materials on the vehicle at a given stop), and average stop spacing (the distance between setting out houses).

In routes constrained by time, route size is sensitive to set-out rate, e.g., the same vehicle on a route with a 20% set-out rate could serve more than twice as many households as on a route with a set-out rate of

80% [16]. Route size was also sensitive to average loading time, e.g., a change in loading time from 10 to 20 s/stop required a reduction in route size of almost 40%. Route size was somewhat less dependent on average stop spacing, with an increase in average stop spacing from 10 to 25 m resulting in an approximately 20% reduction in route size.

Constraining routes by vehicle capacity results in relatively little variance in route time, because the number of stops will be constant [16]. This suggests that trucks of similar capacity, starting routes at similar times, will finish those routes at about the same time, resulting in long lines at unloading locations. When comparing trucks of 15, 20, and 25 yd³ capacity (11.5, 15.3, and 19.1 m³), the medium-sized truck was rarely the lowest-cost option. Large vehicles run up against a time constraint before filling to capacity whereas small vehicles fill well before the end of the collection day. In most cases, either situation was preferred to that of a medium-sized vehicle that meets both constraints at approximately the same point.

Dahlén et al. [17] investigated the effect of different collection systems on recycling in six municipalities in Sweden. PAYT schemes based on weight-reduced residual household waste collection by 50%; however, preliminary results indicated a relatively high rate of impurities (12% in 2004), i.e., wrongly sorted materials, in the biowaste fraction. Curbside collection programs recycled more metal, plastic, and paper packaging than drop-off programs. When separate collection of biodegradables was included in the curbside system, the overall sorting of dry recyclables increased. It appeared that the handling of biodegradables facilitated sorting dry recyclables.

In Sweden, a producer responsibility ordinance mandates that producers, i.e., plastic manufacturers, collect and recycle packaging materials. Hage and Soderholm [18] investigated the relationship between collection rates of household plastic packaging waste and various factors for 252 Swedish municipalities. The results of a regression analysis indicate that local policies, geographic/demographic variables, socioeconomic factors, and environmental preferences provide some explanation of inter-municipality collection rates. The collection rate was positively affected by increases in the unemployment rate, the fraction of private houses, and the presence of immigrants (unless newly arrived) in

the municipality. Distance to recycling industry, urbanization rate, and population density were not significantly related to collection rate. Municipalities with weight-based waste management fees typically experienced higher collection rates than those municipalities in which flat and/or volume-based fees were used.

Gomes et al. [19] simulated the costs of compostables collection using a fixed container system and a transfer station, for a municipality with a population of 28,000 inhabitants. The main goal was to compare three scenarios: traditional, unsorted collection; separate collection of compostables; and separate collection of compostables from urban communities with home composting elsewhere. Input data from 2001 were used, including waste quantities, travel times, work crew composition, crew time shifts, vehicles, and containers. Separate collection of compostables resulted in comparable collection costs to traditional, unsorted collection. When only urban compostables were collected and nonurban compostables were home composted, the costs were lower than traditional, unsorted method. This emphasizes the importance of home composting.

McLeod and Cherrett [20] assessed the effects of three domestic waste collection methods using data from three Hampshire, UK authorities: cooperation between neighboring waste collection authorities and vehicles based at waste disposal sites. Cooperation could reduce vehicle mileage by 5.9% if vehicles were optimally reallocated to depots. Vehicle mileage was reduced by 13.5% when vehicles were based at the two waste disposal sites rather than existing depots. This work identified some basic methods for optimizing waste collection.

Optimization

Numerous optimization studies have been applied to MSW Collection, often identifying significant saving opportunities. According to Tavares et al. [21], the collection of MSW can be more than 70% of the total waste management budget. Fuel costs can be a major portion of the collection cost. Optimization of the routing network used for waste collection and transportation can be used to minimize fuel costs. GIS optimization models can take into account the effects

of road inclination and vehicle weight on fuel consumption. Tavares et al. [21] optimized routes two different ways, minimizing the travel distance or fuel consumption. For the Praia city region (Cape Verde), optimizing for fuel consumption used 8% less fuel than optimizing for route distance. When the model was applied to a waste transport problem, a 12% fuel reduction was obtained. The results indicated the importance of designing routes to minimize uphill climbs to reduce fuel consumption. Reducing fuel consumption will also reduce greenhouse gas (GHG) emissions.

Chang et al. [22] used compromise programming within a GIS framework to optimize waste collection in Kaohsiung city, the second largest city in southern Taiwan. Compromise programming was used to optimize multiple objectives: shortest collection distance, lowest collection cost, and shortest collection time. Optimization was performed on the current management districts and new proposed districts. The model was able to significantly reduce the number of collection points and collection distance and time. The number of collection points was reduced from 1,788 to 617, collection distance was reduced from 80,528 to 51,176 m, and collection time from 1,629 to 1,521 min. Furthermore, the average distances residents had to walk to collection points was also reduced, as was the average service frequency of the collection vehicles.

Kardimas et al. [23] used the ant colony system (ACS) algorithm, within a GIS environment to monitor, simulate, and optimize solid waste collection in a portion of Athens that produced an amount of waste equivalent to the vehicle capacity. The program was able to reduce travel time by approximately 25%, resulting in an overall reduction in collection time (travel plus loading) of approximately 9%. Filipiak et al. [24] used the Chinese postman problem algorithm to optimize MSW collection costs in the Township of Millburn, NJ. They were able to develop a set of optimal collection routes that reduced the distance traveled by approximately 10%. Nuortio et al. [25] optimized vehicle routes and schedules for collecting municipal solid waste in Eastern Finland using a guided variable neighborhood thresholding metaheuristic. The method was able to optimize three routes by an average of 46%.

Kima et al. [26] presented algorithms for collection vehicle routing problems that consider multiple unloading trips per day and driver's lunch breaks. This scenario is most appropriate for commercial waste collection, in which each vehicle can make many unloading trips per day. The algorithms were successfully implemented and used by Waste Management, Inc. In one case study, a collection scenario requiring ten 9-h routes with productivity of 57.06 yd³/h (43.63 m³/h) was improved to nine 9-h routes and 63.40 yd³/h (48.47 m³/h). This reduced the annual number of routes by 984, resulting in a savings of \$18 million.

Green House Gas Emissions and Other Environmental Impacts

The collection of municipal solid waste generates pollution, e.g., vehicle emissions. For isolated communities, the pollution generated by long transport distances required to bring recyclables to processing or manufacturing centers may outweigh the benefits of avoided extraction and processing of virgin materials. A number of researchers have studied the environmental impacts of collection.

Batool and Chuadhry [27] estimated greenhouse gas (GHG) emissions from solid waste management for Data Ganj Bukhsh Town (DGBT) in Lahore, Pakistan using the life cycle assessment (LCA) methodology. There were 1,624,169 people living in 232,024 dwellings in DGBT, with an average per capita MSW generation rate of 0.84 kg/day. Solid waste management options considered included the collection and transportation of waste, composting, biogasification, collection of recyclables with single and mixed material bank container systems, and landfilling. CO₂ equivalents, emitted and avoided, were based on electricity consumption and production, fuel consumption, and emissions.

The baseline scenario was the current MSW system, collection of waste for landfilling without energy recovery [27]. In Scenario 1, 70% of biowaste was collected at the curb for composting. In scenario 2, the biowaste was collected for biogasification. Scenarios 3 and 4 involved drop-off systems for recyclables. Scenario 5 was a combination of scenarios 1 and 3. Scenario 6 combined scenario 5 with energy recovery at the landfill.

LCA results showed that the baseline scenario produced 838,116 tons of CO₂ equivalents annually. Scenarios 1, 2, and 5 resulted in reductions ranging from 58% to 75%. The recycling only options, scenarios 3 and 4, resulted in little savings, primarily because of the low percentage of materials available for recycling (26%). The sixth scenario was a carbon sink, with 33,773 tons stored each year. The authors' calculations indicate that GHG emissions associated with collection are less than reductions associated with recycling.

Solano et al. [28] developed an integrated solid waste management optimization and life cycle assessment (LCA) model that considers cost, energy, and environmental releases. They applied the model to a number of scenarios using a hypothetical, but realistic case study representing an urban region of medium size [29]. Both cost and GHG emission minimization could be determined by the model.

A minimum-cost MSW management system was identified in which residential and multifamily sectors were served by a recyclable materials drop-off center [29]. Approximately 5% of the waste of these two sectors was recovered. The remaining 95% was collected as mixed MSW and disposed of in a landfill. Approximately 26% of commercial sector waste was collected as presorted recyclable material, while the rest was collected as mixed MSW and disposed of in a landfill. Collection and landfill costs constituted 83% of the net cost. The drop-off option for recyclables was incorporated into the minimum-cost MSW system because it costs very little and was easily offset by the revenue generated from the recyclable materials.

According to Solano et al. [29] the majority of MSW management environmental impacts are associated with collection and landfill activities. Collection uses significant amounts of energy, while landfills emit significant amounts of GHG. This assumes that landfill gases are not captured or are flared.

In the minimum GHG emission scenario, the model selected recycling for some combustible and noncombustible items [29]. Recycling of combustible items was favored when recycling was more beneficial than combustion with energy recovery. This was the case for paper and plastic. Recycling of noncombustible items was favored when emissions savings at the remanufacturing facility were greater than the emissions associated with recyclables recovery activities.

This was the case for metal and glass. A costly mixed waste MRF was selected to recover those recyclables not separated by the residents, emphasizing the GHG emissions benefit of recyclable recovery. Almost the entire commercial waste stream was combusted.

Though not adequately described in the paper, it appeared that the only traditional landfilling option considered was one without energy recovery [29]. Other authors have concluded that landfilling with highly efficient gas capture and energy recovery is an optimal solution for minimizing GHG emissions. This points to the importance of base assumptions in LCA analyses.

Iriarte et al. [30] estimated the environmental impacts of three collection systems using an LCA model. Estimates were made for a scenario in which 1,500 tons a month of MSW were generated in a European urban locality with a density of 5,000 inhabitants/km². In the multi-container system, citizens placed materials in permanent containers located in two areas of the street. The organic and residual fractions were placed in containers at the curb at a maximum distance of 50 m from each household. Containers for glass, paper, and packaging were located at a maximum distance of 300 m.

In the door-to-door system, citizens left each fraction outside their house, according to a preestablished collection schedule. The organic fraction was placed at the curb in a community bin. Paper, packaging, and residual waste was left in plastic bags. Glass was collected in permanent containers situated at a maximum distance of 300 m from the citizens.

The mobile pneumatic system used a network of pipes and pneumatic suction trucks. Citizens dropped each fraction of waste through doors located in the street into underground tanks. The tanks were emptied into suction truck via a network of underground pipes connected to suction points situated in the street. Organic, paper, packaging, and residual fractions were collected pneumatically at preset times. Glass was stored in the underground containers, but was collected by top loader trucks.

Three stages of the collection system were explored separately: storage; urban transport; and intercity transport [30]. Storage included containers and, for the pneumatic system, the tanks and underground pipes. Urban transport included loading and transport

from collection points to the locality border. The storage and urban transport stages were together called the urban stages. Intercity transport included transporting the materials from the locality borders to the recycling, processing, or disposal facilities and unloading. Collection vehicles were used for urban and intercity transport.

For the urban stages, the multi-container system had the least environmental impact of all systems [30]. The mobile pneumatic system had greatest environmental impact for global warming, fresh water aquatic ecotoxicity, terrestrial ecotoxicity, acidification, and eutrophication. The door-to-door system had the greatest environmental impact for abiotic depletion, ozone layer depletion, and human toxicity. The intercity transport stage was very important. For intercity distance of 11 km and upward, this stage contributed the most global warming impact and energy demand, for all the systems. This points to the importance of using transfer stations to transfer materials to more efficient transport vehicles.

Di Nino and Baetz [31] examined the link between air emissions from vehicles involved in the collection and transport of recyclable and residual materials and the urban form, i.e., population density, land-use mix, land-use pattern, and scale of development. Two hypothetical cities were evaluated, representing extremes of urban form. The “spread” city was comprised of large single-family homes and homogeneous land-use development with 6.3 units/ha. The “Nodal” city contained compact developments at higher densities and intermixed land use, with 18.01 units/ha. As expected, the “spread” city resulted in significantly more air emissions associated with recyclables and residuals collection. This was true for all of the air pollutants evaluated, including hydrocarbons, carbon monoxide, carbon dioxide, oxides of nitrogen, and particulates. For many of the pollutants, the “Spread” City resulted in more than twice the emissions of the “Nodal” city.

Special Collection Programs

Collection programs for household hazardous waste are common. Though some programs employ curbside collection, drop-off centers are much more common. In some cases, the drop-off centers are not active all the time. Some drop-off centers only collect materials once



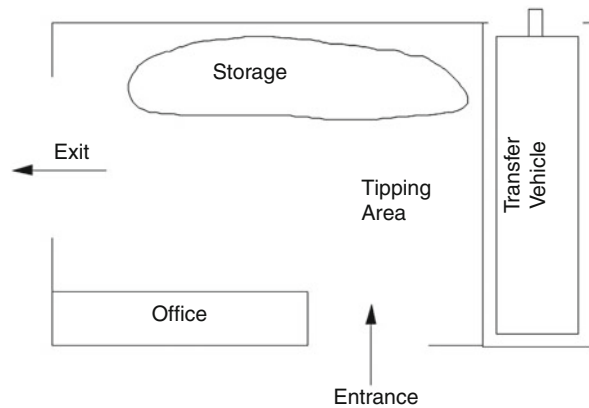
Waste Collection and Transport. Figure 3
Battery collection box, Sweden

a month or four times a year. Common materials collected are paints, pesticides, cleansers, oil, electronics, and batteries.

Some communities have separate collection programs for batteries [32]. In some jurisdictions, only household batteries are considered a non-hazardous waste. In such cases, non-household batteries are often excluded to avoid hazardous waste regulations. Collected batteries can be disposed of at hazardous waste facilities or recycled. Collection programs range from drop off to curbside collection programs. Some drop-off programs have used widely distributed battery collection boxes (Fig. 3). Curbside collection can be with non-hazardous recyclables or hazardous waste. Program costs can be high, especially for curbside collection programs. In Spokane, WA, curbside collection with recycling cost \$2,260 per ton, while curbside collection with landfill disposal cost \$1,822 per ton [32].

Transfer and Transport

Transfer stations are an important part of municipal solid waste management systems. The location of transfer stations has traditionally been driven by a desire to minimize transport costs. Transfer stations are used to combine many small loads from collection vehicles into a few large loads in transport vehicles.



Waste Collection and Transport. Figure 4
Simple transfer station schematic

Transfer stations are typically located close to population centers because it is cheaper to transport large amounts of waste over long distances in large loads than in small ones. Material recovery facilities are a special type of transfer station, used to process recyclables or compostables before transferring them to long-haul vehicles.

A simple transfer station is shown in Fig. 4. Not shown in the figure is a scale, commonly used to determine the amount of material brought by each vehicle. Collection vehicles empty in the tipping area. This can be onto the tipping floor, into a pit, or directly onto a transfer mechanism. Transfer mechanisms include front-end loader, conveyor, and crane. The transport vehicle is often located at a lower level than the tipping floor, allowing transfer by pushing material horizontally.

Wilson and Vincent [33] used on-board global positioning system (GPS) recorders to collect field data on the movement of solid waste collection vehicles at transfer stations. Five vehicles using four transfer stations were compared for 1 year. Spatial data were analyzed using geofences, determining the time each truck spent on four activities: queuing at the scale, weighing, queuing at the tipping floor, and unloading waste. The authors were able to identify and measure queuing delays. The average total time was 16.4 min per visit, with a standard deviation of 14.3 min. Time per visit ranged from 2 to 111 min per visit. The distribution was positively skewed. Multi-compartment vehicles

(co-collection and recycling trucks) spent more time at the unloading facilities than single compartment vehicles. At one facility, the longest delays were encountered queuing at the scale, at two facilities trucks experienced delays queuing at the tipping floor, while at the fourth facility no significant delays developed.

Collection vehicles are designed for collection, while transport vehicles are designed for long-distance hauling. If the final destination for the collected materials is far away, costs and environmental impacts can be reduced by transferring materials to more efficient transport vehicles, which include tractor trailers, trains, and ships.

Optimization

Numerous authors have developed algorithms for optimizing transfer and transport. Komilis [34] presented two conceptual mixed integer linear optimization models to optimize the transfer and transport of municipal solid waste (MSW) prior to landfilling. One model was based on minimizing time. The other was based on minimizing total cost. Both models identified an optimum route to haul MSW from generation points to landfills via transfer stations. The models assumed the locations of generation, transfer, and landfills were fixed. The basic input data were distances, average vehicle speeds, haul costs, equipment and facilities' operating and investment cost, labor cost, and tipping fees.

A case study at the Municipality of Athens was used to illustrate the models [34]. The time-based optimization model is simpler, as it is primarily based on the distances among generation points, transfer stations, and landfills; however, it is best used when no transfer stations are included in the system, as it is difficult to interpret the effect of the average speeds of the collection and transport vehicles. The cost optimization model is more appropriate for systems with transfer stations. For the baseline scenario – 100 tons/day MSW, transfer station location 15 km from generation point, and queue times of 0.5 h – the use of the transfer station was warranted when the distance to the landfill was at least 65 km. This is referred to as the threshold distance. Increasing the tons/day to 500 reduced the threshold distance to 56 km. The minimum threshold distance,

for any amount of waste, was 54 km. Increasing the queuing time at the landfill to 1 h decreased the threshold distance to 56 km for the baseline scenario. Similarly, moving the transfer station location to 5 km reduced the threshold distance to 48 km.

Green House Gas Emissions and Other Environmental Impacts

Transfer stations can negatively impact the environment. Bovea et al. [35] compared the environmental impact of taking municipal wastes directly to the nearest waste treatment facility with waste management systems that include a transfer station. They applied a life cycle methodology to the Plana region of Castellón, Spain. They found that reductions in environmental impact from incorporating a transfer station ranged from 14.0 to 20.7%, depending upon the LCA method used.

Eshet et al. [36] estimated the economic value of externalities related to Israeli waste transfer stations. Most externalities were associated with impacts on residents living near transfer stations, such as noise, odor, litter, vermin, visual intrusion, and perceived discomfort. After mapping active transfer stations in Israel, problem sites near residential areas were identified and four were selected for further study. The Hedonic Price Method was used to examine the impact on property values. The maximum spatial impact occurred within 2.8 km of a transfer station. There was an increase of approximately \$5,000 in housing price for each additional kilometer away from the site. There also was a 0.06% rise in the price of the average house for each 1% increase in average distance from the local transfer station. Transfer stations create externalities that should be taken into account during the siting process and when considering compensation for nearby residents.

Eisted et al. [37] reviewed electricity and fuel consumption associated with the collection, transfer, and transport of waste and quantified the associated greenhouse gas (GHG) emissions. Emission factors were assigned a global warming potential and aggregated into global warming factors (GWFs) that can be used to estimate the contribution to global warming from collecting, transporting, and transferring 1 metric ton of wet waste. Not considered were GHG emissions

Waste Collection and Transport. Table 2 Global warming factors (GWFs) for six scenarios

Scenario	Collection	Transfer	Transport	Total
Residual waste with 20 km of transport by truck, urban apartment blocks with pneumatic system	43–46	–	4–11	47–57
Residual waste with 20 km of transport by truck, urban apartment blocks with dumpsters	5.0–5.5	–	4–11	9–17
Residual waste with 150 km of transport by truck, single-family homes, transfer station	10–11	0.05–4.5	14–28	24–44
Recyclable paper with 2,000 km of transport by truck, drop-off centers, transfer station	6–11	0.05–4.5	182–380	189–396
Recyclable paper with 3,000 km of transport by diesel train, drop-off centers, transfer station	6–11	0.05–4.5	6–174	13–190
Recyclable materials with 10,000 km of transport by ship, private car brings materials to transfer station, 100 km to port	100–300	0.10–8.9	29–59	129–368

related to production, maintenance, and disposal of vehicles, equipment, infrastructure, and buildings. The estimated GWFs varied from 9.4 to 368 kg CO₂-equivalent (kg CO₂-eq.) per metric ton of waste, depending on method of collection, capacity, and choice of transport equipment, and travel distances. Six scenarios are described in Table 2. GHG emissions can be reduced by avoiding the use of private cars and by optimizing long-distance transport, i.e., using trains and ships. As also found by Iriarte et al. [30], the environmental impact of pneumatic collection was very high.

Chunping et al. [38] measured environmental quality at five MSW Transfer Stations in Beijing from 2001 to 2006. Wastewater, air, and noise were monitored during wet and dry seasons. Levels of total suspended particles, odor, noise, ammonium nitrogen, chemical oxygen demand, biochemical oxygen demand, suspended solid, and fecal coliform were all above the criteria limit in China on numerous occasions. Poor environmental quality was measured at each transfer station in at least 5 of the 7 years evaluated.

Modern waste management in Western Europe has resulted in a remarkable increase in waste

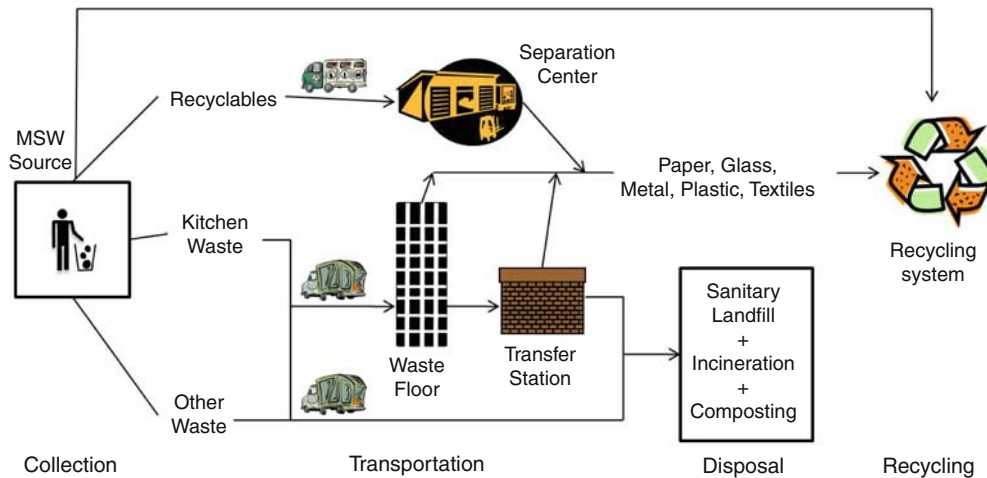
transportation. Salhofer et al. [39] investigated the ecological significance of increased waste transport for recycling refrigerators, paper, polyethylene films, and expanded polystyrene. Specifically, they examined the point at which increasing transportation distance drove the overall environmental impacts associated with recycling over that of disposal. Only in the case of expanded polystyrene were disposal impacts lower than recycling within the upper range of practical transportation distances. For other materials, disposal impacts were higher for all practical distances, indicating that efforts should focus on increasing recovery, e.g., through curbside recycling.

Case Studies

Two case studies further point to the importance of collection, transfer, and transport. The first, Beijing, China, describes a modern waste management system with little recycling. The second, Toronto, Canada, has significant recycling.

Current situation: Beijing, China

Beijing had an urban population of approximately 13.33 million in 2006 [40]. The average generation



Waste Collection and Transport. Figure 5
Beijing MSW system. Adapted from Zhen-Shan [40]

rate of MSW in 2006 was 0.85 kg/capita/day. MSW generation in Beijing City increased from 1.04 million tons in 1978 to 4.134 million tons in 2006. Waste was predominately food waste (63.39%), followed by paper (11.07%), plastics (12.7%), and dust (5.78%), with all other wastes (tiles, textiles, glass, metals, and wood) accounting for less than 3%. In 2009, 90% of MSW generated in Beijing was landfilled, 8% incinerated, and 2% composted. Figure 5 is a schematic of the waste management system. Separate collection was available as a demonstration project to approximately 4.7 million people until 2007. Important MSW characteristics included moisture (61%), combustible waste (92%), recyclable waste (25%), compostable waste (77%), and low calorific value (4,600 kJ/kg). Though the fraction of the combustible waste is high, the low calorific value and high moisture content indicate that incineration would be ineffective without fuel addition or diversion of wet materials.

In 2006, Beijing City expended approximately \$243 million on MSW management. Staffing accounted for \$108 million, followed by waste collection-transportation (\$47 million), treatment (\$33 million), street sweeping (\$32 million), equipment maintenance and repair (\$18 million), and other costs (\$6 million). There were six transfer stations in Beijing, where waste was compressed and transferred to more efficient vehicles for transport to landfills or incineration/composting

facilities. At two transfer stations, waste was separated for recycling. There were 13 landfill sites and four incineration/composting plants, with a total designed capacity of 10,350 tons/day. At the end of 2005, there were 1,242 formal recycling sites in Beijing City, including 704 community-based sites and 538 floating sites. Approximately 1.6 million tons of materials were recycled in 2006, 7.3% higher than in 2005, including 0.36 million tons of paper, 0.87 million tons of metal, 0.17 million tons of plastic, and 0.08 million tons of glass. There were approximately 300,000 people living off the proceeds of waste recycling. Materials recycled by waste pickers were worth \$0.44 million. The waste management system is shown in Fig. 5.

Aeration was the most popular method for odor control in separation workshops. Sprinklers are used to control dust. Approximately 50 tons of wastewater (including leachate and sewage) were generated each day at the transfer station with recyclables recovery. Leachate was mainly from waste compression and dust control water. An anaerobic–anoxic–aerobic reactor followed by a membrane filtration was used to treat the leachate.

Asuwei was the oldest and biggest landfill site, with a design capacity of 2,000 tons/day. Approximately 5,000 m³ of landfill gas was generated every day in 2006. Before 2001, landfill gas was discharged. As of 2009, it is used to generate electricity.

Future Simulations: Toronto, Canada

Diaz and Warith [41] applied the Waste Analysis Software Tool for Environmental Decisions (WASTED) model – which provides a comprehensive view of the environmental impacts of MSW management systems – to Toronto, Canada. The model includes waste collection,

material recovery, composting, energy recovery, and landfilling. WASTED uses compensatory systems to account for avoided environmental impacts. Upstream compensation incorporates the effects of processes that precede the waste management system (e.g., raw material extraction), while downstream compensation incorporates environmental effects of processes that

Waste Collection and Transport. Table 3 Waste Analysis Software Tool for Environmental Decisions (WASTED) results for Toronto, 2001 to 2021 [41]

	Collection/transport	Recycle	Compost	Landfill	Total
MSW Tonnes managed	1.956E10 ⁷	5.585E10 ⁶	1.905E10 ⁶	1.207E10 ⁷	1.956E10 ⁷
Energy consumed (GJ)	7.104E10 ⁹	3.641E10 ⁸	2.875E10 ⁶	4.700E10 ⁸	7.941E10 ⁹
Air pollutants					
Total CO ₂ (ton)	6.024E10 ⁵	−2.010E10 ⁷	2.096E10 ⁶	2.446E10 ⁶	−1.495E10 ⁷
Non-biogenic CO ₂ (ton)	6.024E10 ⁵	−2.010E10 ⁷	−3.483E10 ⁵	−1.135E10 ⁶	−2.098E10 ⁷
CH ₄ (ton)	7.246E10 ²	−5.481E10 ³	2.932E10 ^{−1}	3.250E10 ⁵	3.202E10 ⁵
CO ₂ equivalent (ton)	6.177E10 ⁵	−2.021E10 ⁷	2.096E10 ⁶	9.270E10 ⁶	−8.227E10 ⁶
Non-biogenic CO ₂ equivalent (ton)	6.177E10 ⁵	−2.021E10 ⁷	−3.483E10 ⁵	5.689E10 ⁶	−1.425E10 ⁷
NO _x (ton)	4.233E10 ³	−2.983E10 ⁴	1.713	−1.781E10 ³	−2.738E10 ⁴
VOC _s (ton)	1.861E10 ³	3.039E10 ⁴	3.238E10 ³	2.732E10 ³	3.822E10 ⁴
SO _x (ton)	1.147E10 ³	−2.362E10 ⁴	4.643E10 ^{−1}	−4.423E10 ³	−2.690E10 ⁴
PM (ton)	3.360E10 ²	−8.417E10 ³	1.360E10 ^{−1}	−4.485E10 ²	−8.529E10 ³
HCl (ton)	1.113E10 ^{−2}	3.200E10 ⁴	0.000	−2.626E10 ²	3.174E10 ⁴
Pb (ton)	2.631E10 ^{−2}	21.428	1.035E10 ^{−5}	−17.277	41.77
Cd (ton)	2.083E10 ^{−3}	4.074	0.000	−49.136	−45.060
Hg (ton)	8.468E10 ^{−4}	2.574E10 ^{−3}	3.421E10 ^{−7}	−3.623E10 ^{−2}	−3.280E10 ^{−2}
NH ₃ (ton)	0.000	0.000	7.237E10 ²	0.000	7.237E10 ²
Dioxins (kg)	2.676E10 ^{−12}	3.718E10 ^{−9}	3.162E10 ^{−16}	−4.468E10 ^{−8}	−4.096E10 ^{−8}
H ₂ S (ton)	0.000	0.000	0.000	42.731	42.731
Water pollutants					
BOD (kg)	7.956E10 ^{−1}	1.248E10 ³	3.220E10 ^{−4}	2.321E10 ⁶	2.323E10 ⁶
COD (kg)	7.956E10 ^{−1}	1.248E10 ³	3.220E10 ^{−4}	3.316E10 ⁶	3.318E10 ⁶
Pb (kg)	9.472E10 ^{−2}	0.425	3.824E10 ^{−5}	8.414	8.934
Cd (kg)	1.023E10 ^{−2}	−1.748E10 ²	4.139E10 ^{−6}	11.055	−1.637E10 ²
Hg (kg)	9.738E10 ^{−5}	1.054E10 ^{−4}	3.938E10 ^{−8}	0.131	0.131
NH ₃ (ton)	0.000	0.000	0.000	2.211E10 ⁵	2.211E10 ⁵
Dioxin equivalent (ton)	0.000	−6.271E10 ^{−5}	0.000	0.000	−6.271E10 ^{−5}

take place after the waste management (e.g., avoided fossil fuel consumption from energy recovery from waste).

Diaz and Warith [41] applied the WASTED model to the city of Toronto, Canada from 2001 to 2021 (Table 3). From the table, one can see that collection was responsible for a majority of the energy consumption and significant portions of GHG emissions and other gas emissions and water pollutants. The authors also compared the results to the outcomes from two previous models, the Integrated Waste Management Model for Municipalities (IWM), and the Waste Reduction Model (WARM). While the models were in general agreement, fairly large differences were reported, indicating the importance of the underlying assumptions in the LCA models.

Future Directions

Waste collection and transport need to be made more efficient to conserve more natural resources and emit fewer pollutants, including green house gases. One way this can be done is to use the many optimization algorithms already developed, in conjunction with geographical information and global positioning systems, to design more efficient routes. Improvements can also be made to collection and transport vehicles, reducing fuel consumptions and pollution emissions. Greater use of trains and ships for transport, where appropriate, should also be pursued. Waste collection improvements will include the spread of recycling to more communities in more countries, through curbside and drop-off collection programs. This will require more material recovery facilities. As landfills become larger and, on average, more distant from waste generation points the use of efficient transport vehicles becomes more important.

Bibliography

Primary Literature

1. Tchobanoglous G, T Hilary, Vigil S (1993) Integrated solid waste management: engineering principles and management issues. Irwin/Mc-Graw-Hill, New York
2. Belton V, Crowe D, Matthews R, Scott S (1994) A survey of public attitudes to recycling in glassgo (UK). *Waste Manage Res* 12:351–367
3. Valeo C, Baetz B, Tsanis I (1998) Location of recycling depots with GIS. *J Urban Plann Dev* 124(2):93–99
4. Bacha H, Mild A, Natter M, Weber A (2004) Combining socio-demographic and logistic factors to explain the generation and collection of waste paper. *Resour Conserv Recycl* 41:65–73
5. Williams I, Taylor C (2004) Maximising household waste recycling at civic amenity sites in Lancashire, England. *Waste Manage* 24:861–874
6. Wilson B, Baetz B (2001) Modeling municipal solid waste collection systems using derived probability distributions. I: model development. *J Environ Eng* 127(11):1031–1038
7. Everett J (1994) Environmental collective action: residential recycling programs. *J Prof Issues Eng Educ Pract* 120(2):158–176
8. Olson M (1965) The logic of collective action. Harvard University Press, Cambridge, MA
9. USEPA (2003) Pay-as-you-throw: a cooling effect on climate change. US Environmental Protection Agency, EPA 530-F-03-008
10. Everett J, Peirce J (1991) National analysis of recycling material recovery rates. *J Resour Manage Technol* 19(4):148–152
11. Everett J, Peirce J (1993) Curbside recycling in the USA: convenience and mandatory participation. *Waste Manage Res* 11:49–61
12. García-Sánchez I (2008) The performance of Spanish solid waste collection. *Waste Manage Res* 26:327–336
13. Rogers J, Englehardt J, An H, Fleming L (2002) Solid waste collection health and safety risks – survey of municipal solid waste collectors. *J Solid Waste Technol Manage* 28(3):154–160
14. Medina M (2005) Serving the unserved: informal refuse collection in Mexico. *Waste Manage Res* 23:390–397
15. WRAP (2008) Kerbside recycling: indicative costs and performance. WRAP, Banbury, Oxon, UK
16. Wilson B, Baetz B (2001) Modeling municipal solid waste collection systems using service probability distributions. II: extensions and applications. *J Environ Eng* 127(11): 1039–1047
17. Dahlén L, Vukicevic S, Meijer J-E, Lagerkvist A (2007) Comparison of different collection systems for sorted household waste in Sweden. *Waste Manage* 27:1298–1305
18. Hage O, Soderholm P (2008) An econometric analysis of regional differences in household waste collection: the case of plastic packaging waste in Sweden. *Waste Manage* 28:1720–1731
19. Gomes A, Matos M, Carvalho I (2008) Separate collection of the biodegradable fraction of MSW: an economic assessment. *Waste Manage* 28:1711–1719
20. McLeod F, Cherrett T (2008) Quantifying the transport impacts of domestic waste collection strategies. *Waste Manage* 28:2271–2278
21. Tavares G, Zsigraiova Z, Semiao V, Carvalho M (2009) Optimisation of MSW collection routes for minimum fuel consumption using 3D GIS modeling. *Waste Manage* 29:1176–1185
22. Chang Ni-Bin, Lu H, Wei Y (1997) GIS technology for vehicle routing and scheduling in solid waste collection systems. *J Environ Eng* 123(9):901–910

23. Karadimas N, Papatzelou K, Loumos V (2007) Optimal solid waste collection routes identified by the ant colony system algorithm. *Waste Manage Res* 25: 139–147
24. Filipiak K, Abdel-Malek L, Hsieh H-N, Meegoda J (2009) Optimization of municipal solid waste collection system: case study. *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management* 13(3):210–216
25. Nuortio T, Kytöjokib J, Niskaa H, Brösysb O (2006) Improved route planning and scheduling of waste collection and transport. *Expert Syst Appl* 30:223–232
26. Kima B-I, Kimb S, Sahoob S (2006) Waste collection vehicle routing problem with time windows. *Comput Oper Res* 33:3624–3642
27. Batool S, Chuadhry M (2009) The impact of municipal solid waste treatment methods on greenhouse gas emissions in Lahore, Pakistan. *Waste Manage* 29:63–69
28. Solano E, Ranjitan R, Barlaz M, Brill D (2002) Life-cycle-based solid waste management I: model development. *J Environ Eng* 128(10):981–992
29. Solano E, Dumas R, Harrison K, Ranjitan R, Barlaz M, Brill D (2002) Life-cycle-based solid waste management II: illustrative applications. *J Environ Eng* 128(10):981–992
30. Iriarte A, Gabarrell X, Rieradevall J (2009) LCA of selective waste collection systems in dense urban areas. *Waste Manage* 29:903–914
31. Nino D, Tony BB (1996) Environmental linkages between urban form and municipal solid waste management infrastructure. *J Urban Plann Dev* 122(3):83–99
32. Shapek R (1995) Local government household battery collection programs: costs and benefits. *Resour Conserv Recycl* 15:1–19
33. Wilson B, Vincent J (2008) Estimating waste transfer station delays using GPS. *Waste Manage* 28:1742–1750
34. Komilis D (2008) Conceptual modeling to optimize the haul and transfer of municipal solid waste. *Waste Manage* 28:2355–2365
35. Bovea M, Powell J, Gallardo A, Capuz-Rizo S (2007) The role played by environmental factors in the integration of a transfer station in a municipal solid waste management system. *Waste Manage* 27:545–553
36. Eshet T, Baron M, Shechter M, Ayalon O (2007) Measuring externalities of waste transfer stations in Israel using hedonic pricing. *Waste Manage* 27:614–625
37. Eisted R, Larsen A, Christensen T (2009) Collection, transfer and transport of waste: accounting of greenhouse gases and global warming contribution. *Waste Manage Res* 27: 738–745
38. Chunping L, Guoxue L, Yiming L, Yanfu L, Jian H (2008) Environmental monitoring and fuzzy synthetic evaluation of municipal solid waste transfer stations in Beijing in 2001–2006. *J Environ Sci* 20:998–1005
39. Salhofer S, Schneider F, Obersteiner G (2007) The ecological relevance of transport in waste disposal systems in Western Europe. *Waste Manage* 27:547–557
40. Zhen-Shan L, Lei Y, Xiao-Yan Q, Yu-mei S (2009) Municipal solid waste management in Beijing City. *Waste Manage* 29:2596–2599
41. Diaz R, Warith M (2006) Life-cycle assessment of municipal solid wastes: development of the WASTED model. *Waste Manage* 26:886–901

Waste Diversion

SHIRLEY THOMPSON, MARY REZAI, JEFF VALDIVIA
Natural Resources Institute, University of Manitoba,
Winnipeg, MB, Canada

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Hierarchy of Waste Diversion
Source Reduction
Recycling
Biological Treatment: Composting and Anaerobic
Digestion
Organic Waste
Future Directions
Bibliography

Glossary

Waste diversion Directs garbage away from landfills or incinerators through reuse, recycling, composting, or gas production through anaerobic digestion. Waste diversion is a key component of effective and sustainable waste management.

Waste management The process of collecting, processing, and disposing of waste.

Extended producer responsibility (EPR) Entails making manufacturers responsible for the entire life cycle of the products and packaging they produce based on the “polluter pays principle.”

Waste prevention (aka Source reduction) Refers to changes in the design, manufacturing process, purchase, or use of materials or products that reduces their contribution to the waste stream.

Definition of the Subject and Its Importance

Waste diversion or landfill diversion is the process of diverting waste from landfill. Diverting waste from the landfill is done through recycling, composting, burning, compacting, etc., to reduce its threat to human health and the environment, make it more disposable, or capture some of its value [1]. There is a hierarchy of waste diversion that lists diversion methods in order of importance, namely, waste prevention (aka source reduction), preparing waste for reuse, recycling/composting, other recovery methods such as energy recovery from waste and biological transformation of waste.

Introduction

Without doubt, diverting waste provides a resource with a value [1], substituting for virgin resources, while conserving the products' embedded energy, and reducing the waste management problems of declining landfill space, leachate, and landfill gas. Unfortunately, the true value of waste is typically underestimated, since placing a value on improved human health or environmental quality is difficult [1]. Whether the value of waste can be captured will often depend on local circumstances, like the cost of labor, equipment, energy, and land [1]. Effective and timely waste diversion policies are needed to ensure the value is seen and help correct the various problems associated with the waste management sector. Waste diversion helps to resolve many of the problems facing waste management including declining landfill space, leachate, landfill gas, and toxins and gases emitted as a result of incineration.

Hierarchy of Waste Diversion

There is a hierarchy of waste diversion, which lists broad waste management measures in order of their importance [2]. The measures from greatest to least importance are waste prevention (aka source reduction), preparing waste for reuse, recycling/composting, other recovery (such as energy recovery from waste) and biological transformation of waste [2, 3]. The EU Directive on Waste stipulates that the following of the hierarchy should avoid human health hazards and environmental damage [2].

Source Reduction

Source reduction refers to changes in the design, manufacturing process, purchase, or use of materials or products that reduces their contribution to the waste stream [3]. It has been observed that source reduction is the best way of decreasing greenhouse gas emissions from the waste sector and decreasing overall waste generation [3, 4]. Reusing materials or products is considered a kind of source reduction as it frees space in landfills for other waste and prevents the creation of new materials or products to be used in their place [3–5].

Extended producer responsibility (EPR) programs, which are programs designed so that producers bear the financial burden of the disposal of their products, is another type of source reduction technique [3, 6]. This tactic attempts to relieve local governments of the costs of managing certain priority products by forcing manufacturers to internalize the cost of recycling within the product price. EPR promotes that producers (usually brand owners) have the greatest control over product design and marketing and therefore have the greatest ability and responsibility to reduce toxicity and waste. In this way, EPR shifts responsibility for waste from government to private industry, obliging producers, importers, and/or sellers to internalize waste management costs in their product prices and ensuring the sustainable and safe handling of the remains of their products [7].

Recycling

Recycling refers to taking used materials and creating new products in order to prevent the disposal of these products in landfills. Recycling material can include glass, paper, metal, plastic, textiles, and electronics. Because recycled materials can act as substitutes for raw materials in many manufacturing processes, recycling helps to reduce the amount of raw materials extracted and processed [3, 6], which results in several favorable outcomes. First, for instance, since recycling aluminum requires far less energy than extracting and processing virgin aluminum, making aluminum cans from recycled aluminum rather than virgin aluminum reduces emissions by 94% and energy use by 93% [3]. Second, since recycled materials are not sent to

landfills, recycling opens up space in landfills for other waste [3, 5].

Ferrara and Missios [8], for instance, investigated the relationship between recycling policy options and recycling behaviors and came to many of the following important findings [9]:

1. User fees for waste disposal increases recycling intensity; however, user fees may lead to illegal dumping.
2. Weekly recycling has a positive effect on the recycling of glass, aluminum, and toxic chemicals, but has a negligible effect on the recycling of newspaper, plastic bottles, tin cans, and cardboard. This result is consistent with the idea that recycling intensity increases when it occurs concurrently with waste collection, that is, when recycling is more convenient.
3. Offering free units under a use fee program for waste disposal negatively impacts recycling.
4. Limiting the number of bags at the curb has a negligible impact on recycling.
5. Promoting curbside recycling increases the rate of non-curbside recycling (i.e., the recycling of toxic materials).
6. For most materials, education level does not significantly affect recycling intensity, except for university undergraduate and/or postgraduate degrees, which increases the intensity of newspaper, aluminum, tin can, and toxic chemical recycling. Education has a positive impact on the recycling of glass above a high school degree.
7. Recycling generally decreases as income increases for newspaper, plastic, and toxic chemicals. This may occur because as income increases, time is more valuable to the recycler.
8. Home ownership is strongly, and positively, correlated to recycling. This suggests that homeowners are more attached to their community and/or are more concerned about their neighbors' perception of them, causing them to recycle more.

Although these behavioral traits are associated with recyclables rather than organics, these findings should apply to how people would react to the implementation of specific organic waste disposal programs.

Biological Treatment: Composting and Anaerobic Digestion

In addition to reusing materials waste products can go through biological treatment, which is the breaking down of material through the action of microorganisms. There are two types of biological treatments: anaerobic digestion and composting. Materials are broken down to carbon dioxide, water, and biomass. Some materials easily break down, others do not. The environment in which the material is placed determines the speed of breakdown.

Organic Waste

Organic waste is comprised of food/kitchen waste, yard waste, and paper [9, 10]. Organic waste amounts to almost 40% in high- and middle-income countries and roughly 30% in low-income countries by weight of municipal solid waste [10]. This portion of municipal solid waste is largely responsible for leachate and landfill gas problems [11]. In landfills, organic wastes, including food waste and yard waste, undergo anaerobic decomposition to produce various gases, including methane [3, 11, 12]. Landfill gas capture for energy recovery is used in many places, since methane is an excellent source of energy when combusted [3, 13]. Methane capture for energy recovery is useful because, although carbon dioxide is produced by the combustion of methane, carbon dioxide is a much less potent greenhouse gas than methane [3, 6, 11, 12].

Therefore, preventing or limiting organic waste from entering landfills would help to solve these problems [10]. There are several benefits to composting the organic portion of the waste stream. According to Gautam et al. [10], these benefits include the following:

1. Backyard and midsize composting at the source reduces the amount of waste to be collected and transported to landfills.
2. Composting reduces the production of leachate and landfill gas, which are both harmful to the environment.
3. Composting increases the life of landfills.
4. Composting and recycling result in a relatively inert waste stream going into landfills so that landfills require less daily cover, less equipment and labor

during operation, and reduced monitoring after closure.

5. Composting produces a useful soil conditioner with some fertilizer value that can often be sold.
6. Composting is one of the least expensive methods of dealing with organic waste.
7. Composting can significantly help municipalities achieve waste diversion targets.

It is probably impossible to draw general conclusions regarding the economics of waste management options [14]. Chynoweth et al. (1992) propose that, for the organic fraction of MSW, anaerobic digestion is better than aerobic composting because aerobic composting requires aeration or mixing, and anaerobic digestion produces a valuable fuel gas, in addition to compost [15]. Braber (1995) found that the anaerobic digestion of the organic fraction of MSW is slightly more expensive than aerobic composting [14]. In Indonesia, Aye and Widjaya (2006) estimate that composting in a centralized facility (CPC) is more cost-effective and environmentally friendly than composting in small, labor-intensive local facilities (CPL), producing biogas and compost simultaneously (BGP), engineering landfills for landfill gas capture to produce electricity (LFE), and business as usual (open dumping). Aye and Widjaya (2006) found that the benefit–cost ratios for CPC, CPL, BGP, and LFE were 2.2, 1.4, 1.3, and 1.03, respectively [16]. In contrast, Renkow and Rubin (1998) found from a survey of MSW composting facilities that their operating costs are generally around \$50 US per ton and that the majority of facilities receive no revenue from their compost [17].

Future Directions

The more widely waste diversion options are available (to residents in single-family households and multifamily buildings and to the institutional, commercial, and industrial sectors) the easier it is to reach a high diversion rate. For this, policy and legislation are essential that support waste diversion efforts. Municipal bylaws limiting waste disposal, charging for garbage collection, or banning divertible materials from landfills can help enforce waste diversion practices. Broader policies, such as

formal state, provincial, or municipal waste diversion targets, can also motivate change.

A number of trends in waste diversion are emerging including zero waste commitments, economic instruments, and green procurement. Zero waste is a long-term commitment to reduce waste through measures such as extended producer responsibility programs, economic instruments to encourage waste reduction, green procurement, and product design that includes end-of-life management. Economic instruments include Pay-As-You-Throw (PAYT) programs that require residents to pay on a volume basis for the disposal of their waste. The program may employ a tag or bag system or a cart system. Other economic instruments include tipping fee surcharges and fines on contaminated loads (waste that has not been source separated). Green procurement and consumer education policies encourage businesses and residents to reuse more and purchase goods with less packaging. For example, some communities or states have introduced bans on organic waste and different products at the landfill and levies on packaging and bottles. This approach requires strong education and enforcement to ensure a level playing field for businesses and change consumer behavior.

Bibliography

Primary Literature

1. Beede DN, Bloom DE (1995) The economics of municipal solid waste. *World Bank Res Obs* 10(2):113–150
2. Europa (2009a) Directive on waste. <http://europa.eu/legislation_summaries/environment/waste_management/ev0010_en.htm#KEY>. Accessed 26 Mar 2010
3. Mohareb AK, Warith M, Narbaitz RM (2004) Strategies for the municipal solid waste sector to assist Canada in meeting its Kyoto Protocol commitments. *Environ Rev* 12:71–95
4. Min H, Galle WP (1997) Green purchasing strategies: trends and implication. *J Supply Chain Manage* 33(3): 10–17
5. Huhtala A (1997) A post-consumer waste management model for determining optimal levels of recycling and landfilling. *Environ Resour Econ* 10:301–314
6. Statistics Canada (2005a) Human activity and the environment: solid waste in Canada. Government of Canada. Catalogue no. 16-201-XIE. <<http://www.statcan.gc.ca/pub/16-201-x/16-201-x2005000-eng.pdf>>. Accessed 15 Mar 2010
7. Nicol S, Thompson S (2007) Policy options to reduce consumer waste to zero: comparing product stewardship and extended

- producer responsibility for refrigerator waste. *Waste Manage Res* 25(3):227–233
8. Ferrara I, Missios P (2005) Recycling and waste diversion effectiveness: evidence from Canada. *Environ Resour Econ* 30:221–238
 9. Thompson S, Sawyer J, Bonam R, Valdivia JE (2009) Building a better waste generation model: validating models with recovery rates at 35 Canadian landfills. *Waste Manage* 29:2085–2091
 10. Gautam S, Bundela PS, Pandey AK, Awasthi MK, Sarsaiya S (2010) Composting of municipal solid waste of Jabalpur City. *Glob J Environ Res* 4(1):43–46
 11. Ackerman F (2000) Waste management and climate change. *Local Environ* 5(2):223–229
 12. Spokas K, Bogner J, Chanton JP, Morcet M, Aran C, Graff C, Moreau-Le Goluan Y, Hebe I (2006) Methane mass balance at three landfill sites: what is the efficiency of capture by gas collection systems? *Waste Manage* 26(5):516–525
 13. Environment Canada (2007) Landfill gas – overview. Government of Canada. <http://www.ec.gc.ca/wmd-dgd/default.asp?lang=En&n=10E36DBA-1>. Accessed 11 Feb 2009
 14. Braber K (1995) Anaerobic digestion of municipal solid waste: a modern waste disposal option on the verge of breakthrough. *Biomass Bioenergy* 9:365–376
 15. Chynoweth DP, Owens J, O'Keefe D, Earle JFK, Bosch G, Legrand R (1992) Sequential batch anaerobic composting of the organic fraction of municipal solid waste. *Water Sci Technol* 25:327–339
 16. Aye L, Widjaya ER (2006) Environmental and economic analyses of waste disposal options for traditional markets in Indonesia. *Waste Manage* 26:1180–1191
 17. Renkow M, Rubin AR (1998) Does municipal solid waste composting make economic sense? *J Environ Manage* 53:339–347
- Canadian Council of Ministers of the Environment (CCME) (2009) Canada-wide action plan for extended producer responsibility. Government of Canada. Retrieved 29 June 2010 from <http://www.ccme.ca/assets/pdf/epr_cap.pdf>
- Christensen TH, Kjeldsen P, Bjerg PL, Jensen BD, Christensen JB, Baun A, Albrechtsen HJ, Heron G (2001) Biogeochemistry of landfill leachate plumes. *Appl Geochem* 16:659–718
- City of Vancouver (2009) Vancouver landfill & vancouver south transfer station rates. Engineering services. Retrieved 20 February 2009 from <http://vancouver.ca/engsvcs/solidwaste/landfill/rates.htm>
- City of Winnipeg (2008a) Water and waste. Retrieved 20 June 2008 from <http://www.winnipeg.ca/waterandwaste/recycle/recyclingReport.stm>
- City of Winnipeg (2008b) Water and waste. Retrieved 20 June 2008 from <http://www.winnipeg.ca/waterandwaste/garbage/residentialReport.stm>
- City of Winnipeg (2009) Brady road landfill. Retrieved 20 February 2009 from <http://www.winnipeg.ca/WaterandWaste/garbage/bradyroad.stm>
- Curmally A (2004) Exploring economic incentives for effective solid waste management in Rhode Island. Retrieved 19 January 2009 from http://envstudies.brown.edu/oldsite/env/theses/master0304/Curmally_Atiyah%20Curmally.pdf
- Czepiel P, Shorter J, Harriss R, Kolb C, Lamb B (2002) The influence of atmospheric pressure on whole-landfill methane emissions observed during a multi-season study. Scientific poster. Retrieved 7 July 2008 from <http://lst.sb.luth.se/iclrs/web/post2002/ppt/poster/Czepiel.pdf>
- Danish Environmental Protection Agency (1999a) Waste 21: waste management plan 1998–2004. Government of Denmark. Retrieved 13 May 2010 from <http://www2.mst.dk/Udgiv/publications/1999/87-7909-571-2/pdf/87-7909-570-4.pdf>
- Danish Government (2007) Greenhouse gas emissions trends and projections in Europe 2007: Denmark – country profile 2007. European Environment Agency. Retrieved 10 June 2010 from http://www.eea.europa.eu/publications/eea_report_2007_5/Denmark.pdf
- David A (2007) Overview of municipal solid waste (MSW) management in Canada. Environment Canada. Government of Canada. In: Proceedings “Municipal waste management workshop: different approaches and solutions.” Retrieved 16 January 2009 from http://www.awma-ovc.ca/AWMA_OVC_WMWorkshop_Alain_David.pdf
- Denman KL, Brasseur G, Chidthaisong A, Ciais P, Cox PM, Dickinson RE, Hauglustaine D, Heinze C, Holland E, Jacob D, Lohmann U, Ramachandran S, da Silva Dias PL, Wofsy SC, Zhang X (2007) Couplings between changes in the climate system and biogeochemistry. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) *Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge/New York

Books and Reviews

- Abbasi PA, Al-Dahmani J, Sahin F, Hoitink HAJ, Miller SA (2002) Effect of compost amendments on disease severity and yield of tomato in conventional and organic production systems. *Plant Dis* 86:156–161
- Anand S, Sen A (2000) Human development and economic sustainability. *World Dev* 28(12):2029–2049
- Barlaz MA, Green G, Chanton JP, Goldsmith CD, Hater GR (2004) Evaluation of a biologically active cover for mitigation of landfill gas emissions. *Environ Sci Technol* 38:4891–4899
- Barrieu P, Sinclair-Desgagné B (2006) On precautionary policies. *Manage Sci* 52(8):1145–1154
- Bishop RC (1993) Economic efficiency, sustainability, and biodiversity. *Ambio* 22(2/3):69–73
- Bogner J, Matthews E (2003) Global methane emissions from landfills: new methodology and annual estimates 1980–1996. *Glob Biogeochem Cycles* 17(2):34–1–34–18
- Borjesson G, Samuelsson B, Galle J, Svensson BH (2000) Methane emissions from landfills: options for measurement and control. Warwickshire, Stratford-upon-Avon

- Department of Environment, Climate Change, and Water (DECCW) (2007) NSW waste avoidance and resource recovery strategy 2007. NSW Government. Retrieved 25 February 2010 from http://www.environment.nsw.gov.au/resources/warr/07226_WARRreport07.pdf
- Department of Environment, Climate Change, and Water (DECCW) (2008) NSW resource recovery industries survey 2006–07: organics reprocessing. NSW Government. Retrieved 1 March 2010 from <http://www.environment.nsw.gov.au/resources/sustainbus/08584organicsRecovery.pdf>
- Department of Environment, Climate Change, and Water (DECCW) (2009a) About POEO legislation. NSW Government. Retrieved 25 February 2010 from <http://www.environment.nsw.gov.au/legislation/aboutpoeo.htm#major>
- Department of Environment, Climate Change, and Water (DECCW) (2009b) DECCW act summaries. NSW Government. Retrieved 25 February 2010 from <http://www.environment.nsw.gov.au/legislation/DECCActssummaries.htm#waarra>
- Department of Environment, Climate Change, and Water (DECCW) (2009c) NSW waste avoidance and resource recovery progress report 2008. NSW Government. Retrieved 26 February 2010 from <http://www.environment.nsw.gov.au/resources/warr/09155WARRreport08.pdf>
- Department of Environment, Climate Change, and Water (DECCW) (2010) The waste and environment levy. NSW Government. Retrieved 3 March 2010 from <http://www.environment.nsw.gov.au/wr/index.htm>
- Department of Justice (2010) Constitution acts, 1867 to 1982. Government of Canada. Retrieved 5 March 2010 from http://laws-lois.justice.gc.ca/eng/Const/Const_index.html
- Department of Local Government (2009) Comparative information on NSW local government councils 2007/08. Retrieved 2 March 2010 from http://www.dlg.nsw.gov.au/dlg/dlghome/documents/Comparatives/Comparatives_2007_08.pdf
- Department of the Environment, Water, Heritage, and the Arts (DEWHA) (2007a) National strategy for ecologically sustainable development. Australian Government. Retrieved 24 February 2010 from <http://www.environment.gov.au/esd/national/nsesd/index.html>
- Department of the Environment, Water, Heritage, and the Arts (DEWHA) (2007b) National strategy for ecologically sustainable development. Australian Government. Retrieved 24 February 2010 from <http://www.environment.gov.au/esd/national/nsesd/strategy/intro.html#GoalsEtc>
- Department of the Environment, Water, Heritage, and the Arts (DEWHA) (2007c) National strategy for ecologically sustainable development. Australian Government. Retrieved 24 February 2010 from <http://www.environment.gov.au/esd/national/nsesd/strategy/wastes.html>
- Department of the Environment, Water, Heritage, and the Arts (DEWHA) (2009a) About the EPBA Act. NSW Government. Retrieved 24 February 2010 from <http://www.environment.gov.au/epbc/about/index.html>
- Department of the Environment, Water, Heritage, and the Arts (DEWHA) (2009b) National environmental protection council. NSW Government. Retrieved 9 June 2010 from <http://www.environment.gov.au/about/councils/nepc/index.html>
- Department of the Environment, Water, Heritage, and the Arts (DEWHA) (2009c) What is protected under the EPBA Act? Australian Government. Retrieved 24 February 2010 from <http://www.environment.gov.au/epbc/protect/index.html>
- Division for Sustainable Development (n.d.) U.N. department of economic and social affairs. Retrieved 18 February 2009 from http://www.un.org/esa/dsd/dsd/dsd_index.shtml
- El-Fadel M, Findikakis AN, Leckie JO (1997) Environmental impacts of solid waste landfilling. *J Environ Manage* 50:1–25
- Elliott A (2008) Is composting organic waste spreading? Statistics Canada. Government of Canada. Retrieved 19 January 2009 from <http://www.statcan.gc.ca/pub/16-002-x/2008001/10540-eng.htm#4>
- Environment Canada (2005b) A guide to understanding the Canadian environmental protection act, 1999: section 2: environmental management in Canada. Government of Canada. Retrieved 5 March 2010 from http://www.ec.gc.ca/CEPARRegistry/the_act/guide04/s2.cfm
- Environment Canada (2005c) A guide to understanding the Canadian environmental protection act, 1999: section 3: CEPA 1999 guiding principles. Government of Canada. Retrieved 5 March 2010 from http://www.ec.gc.ca/CEPARRegistry/the_act/guide04/s3.cfm
- Environment Canada (2006) Greenhouse gases. Government of Canada. Retrieved 16 January 2009 from http://www.ec.gc.ca/pdb/ghg/about/gases_e.cfm
- Environment Canada (2009a) Canada's 2006 greenhouse gas inventory. Government of Canada. Retrieved 11 February 2009 from http://www.ec.gc.ca/pdb/ghg/inventory_report/2006/tab_eng.cfm
- Environment Canada (2009b) Municipal solid waste (MSW). Government of Canada. Retrieved 9 March 2010 from <http://www.ec.gc.ca/drgd-wrmd/default.asp?lang=En&n=7623F633-1>
- Environment Canada Weather Office Environment Canada (2006) Canadian climate normals or averages 1971–2000. Retrieved July 2007. http://www.climate.weatheroffice.ec.gc.ca/climate_normals/index_e.html
- Environment Protection and Heritage Council (2009) National waste overview 2009. Australian Government. Retrieved 2 March 2010 from http://www.ephc.gov.au/sites/default/files/WasteMgt_Nat_Waste_Overview_PRINT_ver_200911.pdf
- Europa (2006) Strategy on the prevention and recycling of waste. Retrieved 29 March 2010 from http://europa.eu/legislation_summaries/environment/waste_management/l28168_en.htm
- Europa (2008a) Integrated pollution prevention and control: IPPC directive. Retrieved 30 March 2010 from http://europa.eu/legislation_summaries/environment/waste_management/l28045_en.htm
- Europa (2008b) Waste incineration. Retrieved 29 Mar 2010 from http://europa.eu/legislation_summaries/environment/waste_management/l28072_en.htm

- Europa (2009b) Waste disposal (until the end of 2010). Retrieved Mar 30 from http://europa.eu/legislation_summaries/environment/waste_management/l21197_en.htm
- Europa (n.d.) The 27 member countries of the European Union. Retrieved 22 Mar 2010 from http://europa.eu/about-eu/27-member-countries/index_en.htm
- European Commission (2009a) What are EU decisions? Retrieved 26 March 2010 from http://ec.europa.eu/community_law/introduction/what_decision_en.htm
- European Commission (2009b) What are EU directives? Retrieved 26 March 2010 from http://ec.europa.eu/community_law/introduction/what_directive_en.htm
- European Commission (2009c) What are EU regulations? Retrieved 26 March 2010 from http://ec.europa.eu/community_law/introduction/what_regulation_en.htm
- European Commission (2009d) What is EU law? Retrieved 26 March 2010 from http://ec.europa.eu/community_law/introduction/treaty_en.htm
- European Compost Network (n.d.) Biological waste treatment in Europe – technical and market developments. Retrieved 19 January 2009 from <http://www.compostnetwork.info/index.php?id=10>
- Federation of Canadian Municipalities (FCM) (2004) Solid waste as a resource: guide for sustainable communities. Retrieved 9 March 2010 from http://www.sustainablecommunities.ca/files/Capacity_Building_-_Waste/SW_Guide_Overview.pdf
- Folke C, Carpenter S, Elmqvist T, Gunderson L, Holling CS, Walker B (2002) Resilience and sustainable development: building adaptive capacity in a world of transformations. *Ambio* 31(5):437–440
- Forcing In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge/New York
- Forster P, Ramaswamy V, Artaxo P, Bernsten T, Betts R, Fahey DW, Haywood J, Lean J, Lowe DC, Myhre G, Nganga J, Prinn R, Raga G, Schulz M, Van Dorland R (2007) Changes in atmospheric constituents and in radiative forcing. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge/New York
- Friesen B (2000) Landfill ban stimulates composting programs in Nova Scotia. *BioCycle Mag* 41(3):53
- Gollier C, Jullien B, Treich N (2000) Scientific progress and irreversibility: an economic interpretation of the precautionary principle. *J Public Econ* 75:229–253
- Google Maps (2010a) Retrieved 24 February 2010 from <http://maps.google.com.au/>
- Google Maps (2010b) Retrieved 4 March 2010 from <http://maps.google.ca/>
- Government of Canada (2010) Provinces and territories. Government of Canada. Retrieved 4 March 2010 from <http://www.canada.gc.ca/othergov-autregouv/prov-eng.html>
- Government of Nova Scotia (2008) Nova Scotia population by county. Government of Nova Scotia. Retrieved 16 March 2010 from <http://www.gov.ns.ca/finance/statistics/agency/common/default.asp?p=1&s=1c>
- Green Manitoba (2009) Waste reduction and recycling support (WRARS) program guide. Government of Manitoba. Retrieved 29 June 2010 from [http://www.greenmanitoba.ca/cim/dbf/wrars_guide_december_09_\(v1.1\).pdf?im_id=392&si_id=1001](http://www.greenmanitoba.ca/cim/dbf/wrars_guide_december_09_(v1.1).pdf?im_id=392&si_id=1001)
- Green Manitoba (2010) HHW events calendar for 2010. Government of Manitoba. Retrieved 29 June 2010 from http://www.greenmanitoba.ca/cim/dbf/2010_hhw_events.pdf?im_id=411&si_id=1001
- Green Manitoba (n.d.) E-waste roundup. Government of Manitoba. Retrieved 29 June 2010 from <http://www.greenmanitoba.ca/cim/1001.dhtm>
- Green Manitoba (n.d.) History of waste minimization in Manitoba. Government of Manitoba. Retrieved 23 June 2010 from http://www.greenmanitoba.ca/cim/dbf/history.pdf?im_id=150&si_id=1001
- Hyder Consulting (2008) Waste and recycling in Australia: final report. Retrieved 25 February 2010 from <http://www.environment.gov.au/settlements/waste/publications/pubs/waste-recycling2008.pdf>
- IPCC (2007) Summary for policymakers. In: Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL (eds) Climate change 2007: the physical science basis. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge/New York
- Kjeldsen P, Barlaz MA, Rooker AP, Baun A, Ledin A, Christensen TH (2002) Present and long-term composition of MSW landfill leachate: a review. *Crit Rev Environ Sci Technol* 32(4):297–336
- Lay J, Li Y, Noike T (1996) Effect of moisture content and chemical nature on methane fermentation characteristics of municipal solid wastes. In: Proceedings: Japan society of civil engineers. *J Environ Syst Eng Div* 1:101–108
- Manitoba Association for Resource Recovery Corporation (MARRC) (2010) On drop of used oil makes a difference. Retrieved 29 June 2010 from <http://www.usedoilrecycling.com/en/mb>
- Manitoba Conservation (2010) News release. Government of Manitoba. Retrieved 29 June 2010 from <http://news.gov.mb.ca/news/index.html?archive=&item=8282>
- Manitoba Conservation (N.d.b.) Waste reduction and pollution prevention (WRAPP) fund. Government of Manitoba. Retrieved 29 June 2010 from <http://www.gov.mb.ca/conservation/pollutionprevention/wrapp/wrappfund.html>
- Manitoba Conservation (N.d.b.) Waste reduction and prevention. Government of Manitoba. Retrieved 29 June 2010 from <http://www.gov.mb.ca/conservation/pollutionprevention/waste/index.html>

- Manitoba Local Government (n.d.) Municipalities. Government of Manitoba. Retrieved 17 May 2010 from <http://web5.gov.mb.ca/public/municipalities.aspx>
- McCabe DJ (1976) Progress report of efforts completed during the second year of baled and unbaled solid waste studies. System Technology Corp, Xenia
- National Capital Commission (2008) About Canada's capital. Government of Canada. Retrieved 5 March 2010 from http://www.capcan.ca/bins/ncc_web_content_page.asp?cid=16297-24515&lang=1&bhpc=1
- Nevens F, Reheul D (2003) The application of vegetable, fruit and garden waste (VFG) compost in addition to cattle slurry in a silage maize monoculture: nitrogen availability and use. *Eur J Agron* 19:189–203
- Nova Scotia Environment (2008) Final report on Nova Scotia's 1995 solid waste resource management strategy: not going to waste. Government of Nova Scotia. Retrieved 10 March 1995 from <http://www.gov.ns.ca/nse/waste/docs/SolidWasteStrategyFinalReport1995.pdf>
- Nova Scotia Environment (2009a) Environmental goals and sustainable prosperity act: progress report 2009. Government of Nova Scotia. Retrieved 11 March 2010 from <http://www.gov.ns.ca/nse/egspa/docs/EGSPA.2009.Annual.Report.pdf>
- Nova Scotia Environment (2009b) Renewal of Nova Scotia's solid waste resource management strategy: consultation summary report 2009. Government of Nova Scotia. Retrieved 10 March 2010 from <http://www.gov.ns.ca/nse/waste/docs/SolidWasteStrategy.2009.Renewal.pdf>
- Nova Scotia Environment (2009c) Solid waste-resource management strategy. Government of Nova Scotia. Retrieved 10 March 2009 from <http://www.gov.ns.ca/nse/waste/swrmstrategy.asp>
- Nova Scotia Environment and Labour (2004) Status report 2004 of solid waste-resource management in Nova Scotia. Government of Nova Scotia. Retrieved 11 March 2010 from <http://www.gov.ns.ca/nse/waste/docs/WasteResourceStatus2004.pdf>
- NSW Legislation (2010) Protection of the environment operations act 1997 no. 156. New South Wales Government
- Office of the Auditor General (2007) Audit of the province's management of contaminated sites and landfills. Government of Manitoba
- Okeke C, Armour A (2000) Post-landfill siting perceptions of nearby residents: a case study of Halton landfill. *Appl Geogr* 20:137–154
- Organic Waste Recycling Unit (2002) Soil carbon sequestration utilising recycled organics: a review of the scientific literature. NSW Agriculture. Retrieved 2 March 2010 from http://www.environment.nsw.gov.au/resources/warr/SPD_ORG_0208SoilCarbonSeq.pdf
- Parliament of Australia (2003) Commonwealth of Australia constitution act. Australian Government. Retrieved 24 February 2010 from <http://www.aph.gov.au/SEnate/general/constitution/parScha1.htm>
- Parliamentary Education Office (PEO) (2009) A multi-tiered system: governing Australia. Australian Government. Retrieved 24 February 2010 from <http://www.peo.gov.au/students/cl/multi.html>
- Pinamonti F (1998) Compost mulch effects on soil fertility, nutritional status and performance of grapevine. *Nutr Cycl Agroecosyst* 51:239–248
- Press, Cambridge/New York
- Read AD (1999) Making waste work: making UK national solid waste strategy work at the local scale. *Resour Conserv Recy* 26:259–285
- RRFB Nova Scotia (2003) An assessment of the state of provincial compost processing facilities. Government of Nova Scotia. Retrieved 16 March 2010 from <http://www.rrfb.com/pdfs/An%20Assessment%20of%20the%20State%20of%20Provincial%20Compost%20Processing%20Facilities.pdf>
- RRFB Nova Scotia (2008) Annual report 2008. Government of Nova Scotia. Retrieved 16 March 2010 from http://www.rrfb.com/pdfs/RRFB_AR_2008.pdf
- Sathaye J, Najam A, Cocklin C, Heller T, Lecocq F, Llanes-Regueiro J, Pan J, Petschel-Held G, Rayner S, Robinson J, Schaeffer R, Sokona Y, Swart R, Winkler H (2007) Sustainable development and mitigation. In: Metz B, Davidson OR, Bosch PR, Dave R, Meyer LA (eds) *Climate change 2007: mitigation. Contribution of working group III to the fourth assessment report of the intergovernmental panel on climate change*, Cambridge University Statistics Canada (2010a) 2006 community profiles. Government of Canada. Retrieved 25 June 2010 from <http://www12.statcan.ca/census-recensement/2006/dp-pd/prof/92-591/index.cfm?Lang=E>
- Statistics Canada (2005b) Land and freshwater area, by province and territory. Government of Canada. Retrieved 4 March 2010 from <http://www40.statcan.ca/l01/cst01/phys01-eng.htm>
- Statistics Canada (2008a) Farm population and total population by rural and urban populations, by province (2001 and 2006 census of agriculture and census of population). Government of Canada. Retrieved 4 March 2010 from <http://www40.statcan.ca/l01/cst01/agrc42a-eng.htm>
- Statistics Canada (2008b) Population and dwelling counts, for Canada, provinces and territories, and census subdivisions (municipalities), 2006 and 2001 censuses – 100% data. Government of Canada. Retrieved 25 June 2010 from <http://www12.statcan.ca/english/census06/data/popdwel/Table.cfm?T=302&SR=1&S=3&O=D&RPP=25&PR=46&CMA=0>
- Statistics Canada (2008a) Disposal and diversion of waste, by province and territory, 2004 and 2006 (diverted materials per capita). Government of Canada. Retrieved 4 May 2010 from <http://www40.statcan.ca/l01/cst01/envir32d-eng.htm>
- Statistics Canada (2008b) Waste management industry survey: business and government sectors. Government of Canada. Catalogue no. 16F0023X
- Statistics Canada (2008c) Waste management industry survey: business and government sectors. Government of Canada. Catalogue no. 16F0023X. Retrieved 12 March 2010 from <http://www.statcan.gc.ca/pub/16f0023x/16f0023x2006001-eng.pdf>

- Statistics Canada (2009) Population by year, by province and territory. Government of Canada. Retrieved 4 March 2010 from <http://www40.statcan.ca/l01/cst01/demo02a-eng.htm>
- Statistics Canada (2010b) Waste disposal, by source, by province. Government of Canada. Retrieved 12 March 2010 from <http://www40.statcan.gc.ca/l01/cst01/envir25a-eng.htm>
- Statistics Denmark (2009) Statistical yearbook 2009. Government of Denmark. Retrieved 22 March 2010 from <http://www.dst.dk/pubpdf/14468/sy2009>
- Statistics Denmark (2010a) Population projections by municipality, age and sex. Government of Denmark. Retrieved 22 March 2010 from <http://www.statbank.dk/statbank5a/default.asp?w=1680>
- Statistics Denmark (2010b) Quarterly population (ultimo). Government of Denmark. Retrieved 22 March 2010 from http://www.dst.dk/HomeUK/Statistics/Key_indicators/Population/pop_quarterly.aspx
- Tanapat S (2004) A feasibility study of municipal waste-to-energy management: measurement of landfill gas quality at Brady Road Landfill in Canada. Unpublished master's thesis, University of Manitoba, Winnipeg
- The City of Sydney (2010) Estimated resident population (ERP). Retrieved 24 February 2010 from <http://profile.id.com.au/Default.aspx?id=148&pg=210>
- The Natural Step (n.d.) The four system conditions. Retrieved 6 March 2009 from <http://www.thenaturalstep.org/en/canada/the-system-conditions>
- Thompson S, Bonam R Best practices for expanding the life of landfills: results of a 2005 Canadian landfill survey. Unpublished manuscript
- Thompson S, Sawyer J, Bonam R, Smith S (2006) Recommendations for improving the Canadian methane generation model for landfills. University of Manitoba prepared for environment Canada, Mar 2006
- Thompson S, Sawyer J, Valdivia JE (2009) Building a better methane generation model: validating models with methane recovery rates from 35 Canadian landfills. *Waste Manage* 29(7):2085–2091
- Tiehm A, Nickel K, Neis U (1997) The use of ultrasound to accelerate the anaerobic digestion of sewage sludge. *Water Sci Technol* 36(11):121–128
- Tire Stewardship Manitoba (2009) 2009 annual report. Retrieved 29 June 2010 from http://www.tirestewardshipmb.ca/cim/dbf/tsm-2009annualreport_web_s.pdf?im_id=54&si_id=1005
- Tuomela M, Vikman M, Hatakka A, Itävaara M (2000) Biodegradation of lignin in a compost environment: a review. *Bioresour Technol* 72:169–183
- US EPA (2009) Municipal solid waste generation, recycling, and disposal in the United States: facts and figures for 2008. EPA 530-F-009-021. US Government. Retrieved 10 May 2010 from http://www.epa.gov/osw/nonhaz/municipal/pubs/msw2008_rpt.pdf
- US EPA (2010a) Inventory of U.S. greenhouse gas emissions and sinks: 1990–2008. US Government. EPA 430-R-10-006. Retrieved 10 May 2010 from http://www.epa.gov/climatechange/emissions/downloads10/US-GHG-Inventory-2010_ExecutiveSummary.pdf
- US EPA (2010b) Methane. US Government. Retrieved 10 May 2010 from <http://www.epa.gov/methane/sources.html>
- van der Werf P, Cant M (2007) State of composting in Canada, part 2. *Solid waste magazine* Dec/Jan:43–46
- Visit Denmark (2009) Geography. Government of Denmark. Retrieved 22 March 2010 from <http://www.visitdenmark.com/usa/en-us/menu/turist/turistinformation/fakta-az/geography.htm>
- Vital Statistics (2006) Annual report. Nova Scotia Government. Retrieved 4 March 2010 from <http://www.gov.ns.ca/snsmr/pdf/ans-vstat-2006-annual-report.pdf>
- Wagner T, Arnold P (2008) A new model for solid waste management: an analysis of the Nova Scotia MSW strategy. *J Cleaner Prod* 16:410–421
- WCED (1987) Our common future: world commission on environment and development. Oxford University Press, Oxford
- Westerman PW, Bicudo JR (2005) Management considerations for organic waste use in agriculture. *Bioresour Technol* 96:215–221
- Wilson EJ, McDougall FR, Willmore J (2001) Euro-trash: searching Europe for a more sustainable approach to waste management. *Resour Conserv Recy* 31:327–346

Waste Management for Sustainable Society

BRIAN BAHOR, MICHAEL VAN BRUNT

Covanta Energy Corporation, Morristown, NJ, USA

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Greenhouse Gas Emissions
 Emissions
 Land Management
 Summary
 Bibliography

Glossary

Energy from waste (EfW) Synonym of WTE used by some companies.

Greenhouse gases (GHG) Carbon dioxide, methane, and other gases contributing to the greenhouse effect.

Integrated waste management Using a combination of all possible means of waste management so as to minimize adverse environmental impacts and optimize resource conservation.

Waste-to-energy (WTE) Recovery of the chemical energy contained in used materials by means of thermal treatment.

Definition of the Subject

The goal of this entry is to establish how an integrated waste management system can be part of the solution to the emerging crises of climate change and inadequate supply of renewable energy for future generations.

Developed and emerging countries societies are consuming natural resources at an increasing rate due to consumer habits and growing population. There is international agreement that a long-term solution to the energy and climate change issues must be a low, or neutral, carbon energy supply. There is no agreement on a solution; however, most discussion focuses on solar or wind energy and even these options are opposed sometimes at a local level. Waste is a resource that, when managed correctly with commercially proven technologies, can also be a sustainable

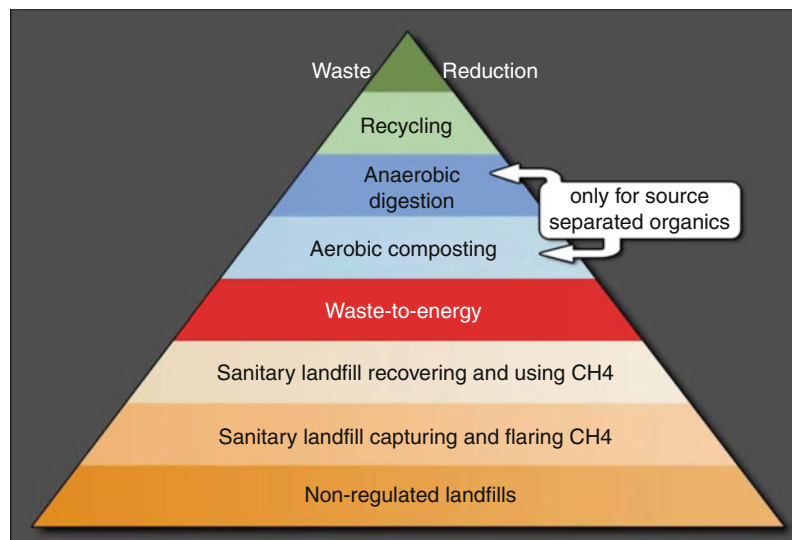
source of renewable energy and contribute to greenhouse gas mitigation.

The waste hierarchy adopted by the European Union [1] and the US Environmental Protection Agency (USEPA) [2] provides a practical guideline for understanding general waste management priorities. Information provided in this entry quantifies the value of an integrated waste management system comprised of technology solutions including recycling, recovery, and disposal.

Introduction

The term “integrated waste management system” refers to a system that includes a combination of management options that provide a stable long-term solution based on a life-cycle assessment. The general principle is to manage the quantity of waste in accordance with the waste hierarchy as illustrated by Fig. 1.

The general concept of the hierarchy is that the “most sustainable” option is at the top because it conserves natural resources, saves or recovers energy, and reduces the emission of greenhouse gases. Options lower in the table are considered to be less desirable because, generally, they are not as effective.



Waste Management for Sustainable Society. Figure 1

The hierarchy of waste management [3]

Waste reduction and reuse are correctly identified as the most preferable solution. However, implementation of these options requires a policy initiative whereas the last three options (recycling, energy recovery, and disposal in landfill) are technology options for the waste remaining after waste reduction and reuse efforts have been implemented. This sustainability evaluation therefore focuses on the technology options.

In regards to sustainability, the standard reference [4] is from the Brundtland commission where sustainable development is defined as development that meets the needs of the present without compromising the ability of future generations to meet their own needs. While that definition is a simple and effective metric, it does not capture the larger concept used by leading sustainability organizations that address a broad range of sustainability challenges, such as global climate change, dwindling resources, water scarcity, and the lack of available low carbon power options that are acceptable to society.

Principles Versus Practice

National and international organizations have determined that waste is a resource that when managed properly can provide a wide range of benefits to society, including preservation of natural resources, clean energy, reduction of greenhouse gas emissions, and stable long-term jobs that provide a community with a stable economy. The best example of a regulatory framework that considers waste as a resource is European Union Directive 2008/98 that establishes the waste hierarchy as a function of their Sixth Environmental Programme [5]. USEPA's 2009 report also determined that proper materials management can provide greenhouse gas

mitigation while also yielding multiple environmental, human health, and economic benefits for communities and the nation; however this position has not been implemented as a matter of policy, legislation, or regulation.

Despite the abundant and growing analysis that supports management of MSW as a resource, and the progress made by the European Union (EU), the practices of many countries lag behind. Table 1 presents the amount of MSW managed by recycling, energy recovery, and landfills in three major areas of the world; North America, the European Union, and East Asia which is primarily China. These areas were selected due to readily available information on each and because each area is unique relative to waste management strategies.

Europe is the only area moving toward an integrated waste management system with Germany leading the effort with over 30 years of legislative efforts to change their waste management system. North American values are largely driven by the USA where there has not been a policy or regulatory driver to change the strong dependence on landfills. East Asia is interesting for several reasons. China is one of the world's largest economies and has adopted a policy to divert MSW from landfills by providing waste-to-energy capacity. China also influences the demand for recycled commodities, with both North America and Europe sending a majority of separated materials to China for transformation to recycled products.

The potential to diversify waste management on a national basis is demonstrated by several European Union (EU) Member States. Information from Eurostat [7] has been used to construct 10-year trends that illustrate how different paths toward advanced waste management have been followed. Note that the

Waste Management for Sustainable Society. Table 1 Allocation of MSW for existing practices

Reference information		Tons of MSW ^a			Percent of total		
Region	2004 ^a	Landfill	EfW	Recycle	Landfill	EfW	Recycle
European Union	379	220	46	99	58	12	26
North America	371	252	22	96	68	6	26
East Asia	442	345	53	44	78	12	10

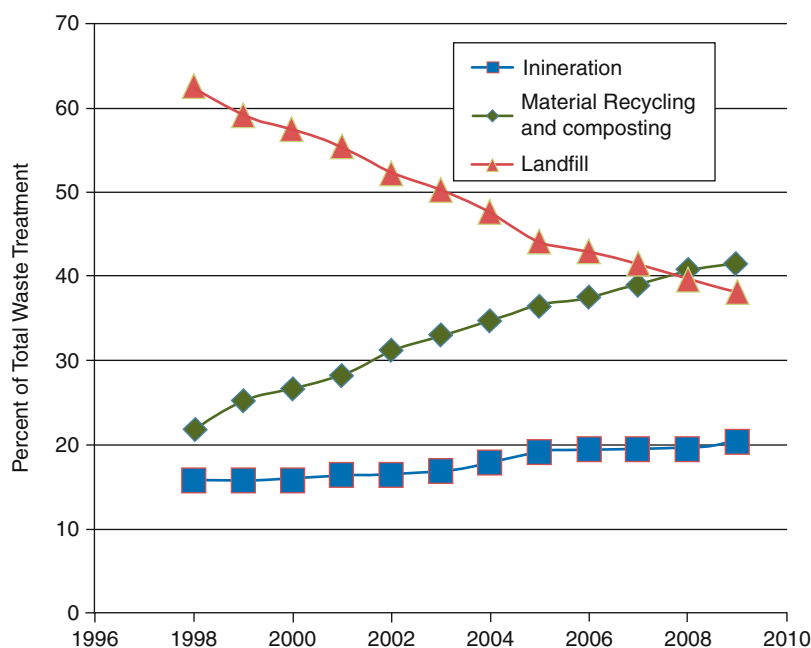
^aMSW values (as million tons) and allocation as % of total is from [6]

Eurostat data is available for several years; however subsequent analysis presented herein is largely based on 2004 information which includes recycling data for conventional commodities (glass, paper, metals, plastic) and compost. It should be noted that waste management in the EU is, to some extent, driven by landfill diversion targets set out in the Landfill Directive. Also, the recent revision to the Waste Framework Directive requires a minimum of 50% overall recycling by 2020.

Figure 2 presents the average results for the 27 Member States and the general trend where MSW is being diverted from landfills to recycling and to a lesser extent to WTE. Figures 3 and 4 present results for Germany and the Netherlands respectively, two countries that have a long-term history of leading Europe in waste management and environmental concerns in general. Figures 5, 6, and 7 demonstrate trends in Spain, Italy, and Greece. Each has a relatively constant % of MSW being managed by WTE with Spain and Italy showing a growth in recycling due to diversion of MSW from landfills. Greece shows limited growth in recycling with no MSW being managed at a WTE facility.

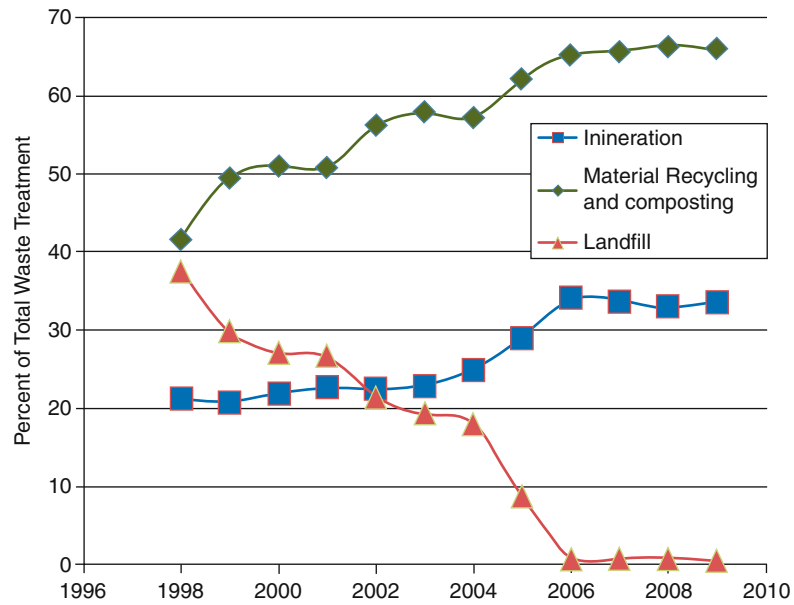
Table 2 presents waste management data for the USA from two different sources. One set is from the USEPA's annual solid waste report [8] with the second set being reported by Biocycle [9, 10]. USEPA results are based on a survey of manufacturing information with waste generation rates being estimated from consumption rates of resources whereas Biocycle data is based on surveys completed by solid waste departments from all 50 states using data reported by waste management operations. The significant difference in results is largely attributable to the large difference in MSW landfilled in the USA, with USEPA estimating 135 million tons in 2008 and Biocycle estimating 265 million tons. Biocycle data is considered to be more useful and accurate and is in fact used by the USEPA in its greenhouse gas inventory report [11]. An independent analysis of USEPA reported values [12] concluded that the actual recycle rate is 27.8% when adjusting values to exclude commercial and industrial materials.

There is insufficient data to track long-term performance of the three waste management options in China.



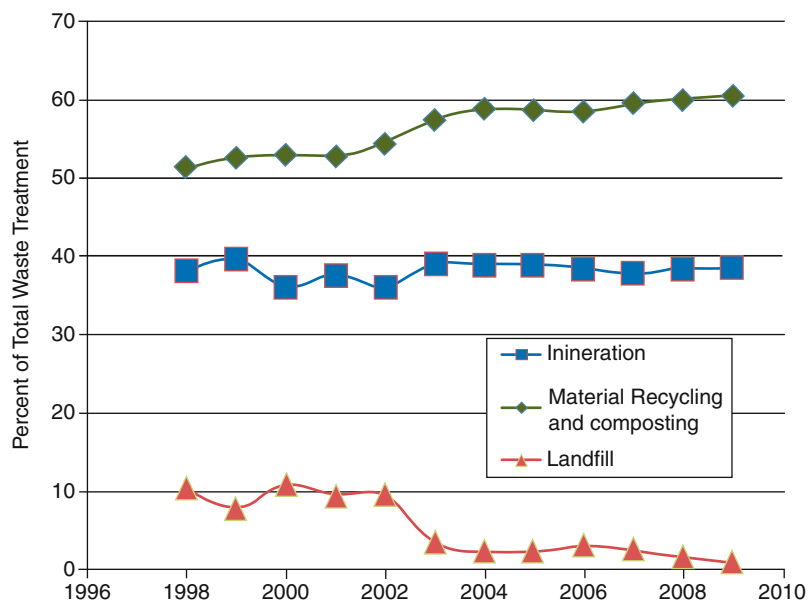
Waste Management for Sustainable Society. Figure 2

Waste management activities: EU 27



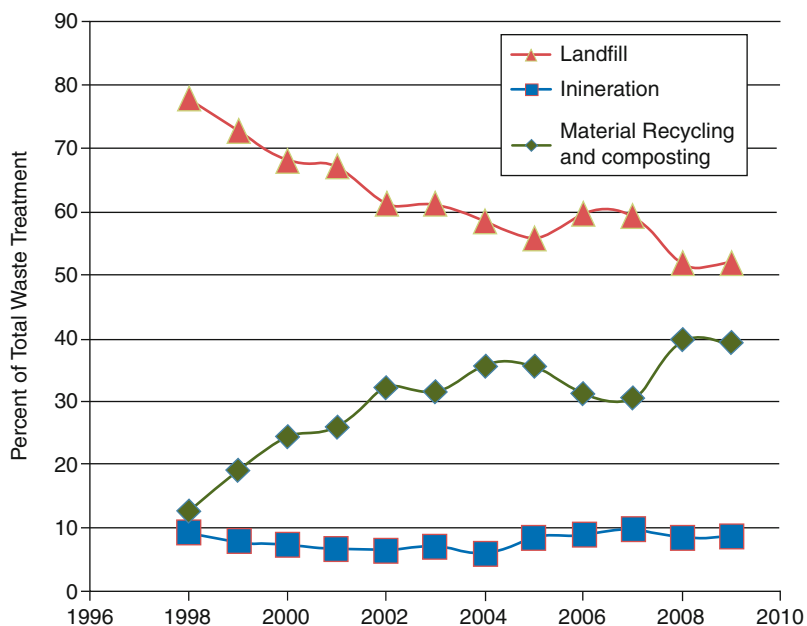
Waste Management for Sustainable Society. Figure 3

Waste management activities: Germany



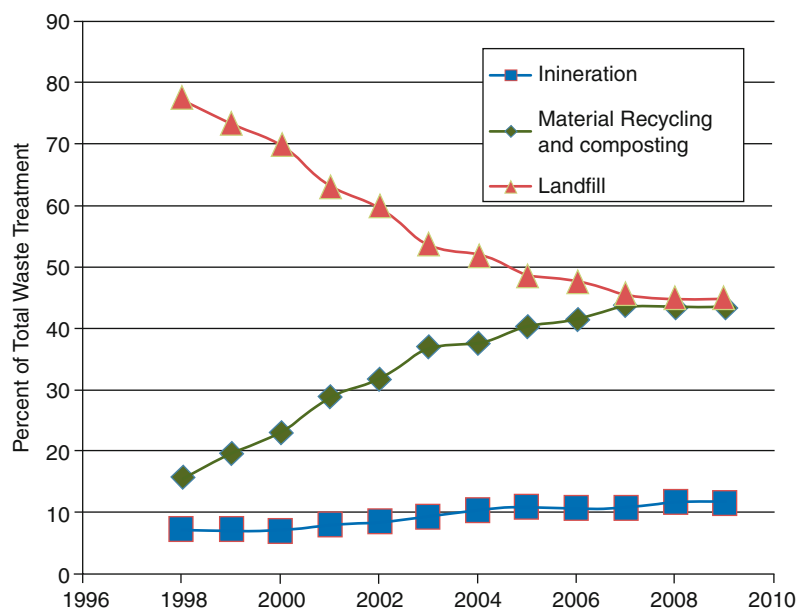
Waste Management for Sustainable Society. Figure 4

Waste management activities: Netherlands



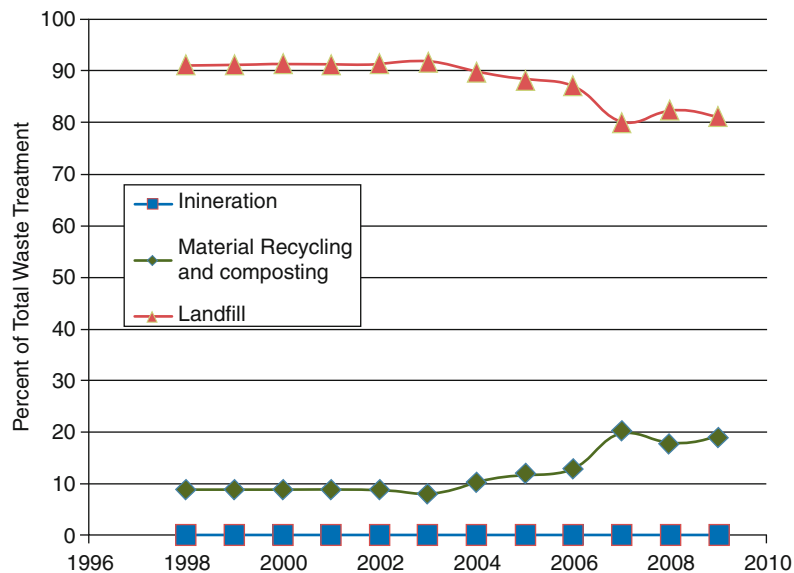
Waste Management for Sustainable Society. Figure 5

Waste management activities: Spain



Waste Management for Sustainable Society. Figure 6

Waste management activities: Italy



Waste Management for Sustainable Society. Figure 7

Waste management activities: Greece

Waste Management for Sustainable Society. Table 2 Waste management in the USA: 2000–2007

Reference		Year ^a		
Source	Option	2000	2005	2007
EPA	Recycling and composting	28.6	31.6	33.3
	WTE	13.9	12.5	12.5
	Landfill	57.5	55.9	54.2
Biocycle	Recycling and composting	32	28.5	28.6
	WTE	7	7.4	6.9
	Landfill	61	64.1	64.5

^aBiocycle data from 2004 are used in 2005 and from 2006 in 2007 to enable a comparison

Observations from long-term Eurostat and USEPA data are:

- The most recent average recycle rate for the EU's 27 member States (41.5%) is significantly greater than the average USEPA adjusted factor (27.8%) and *Biocycle's* (24%) factor.
- The USA continues to landfill the majority of its MSW although it is generally recognized to be the least sustainable alternative according to waste hierarchies.
- A national program, such as the EU's landfill directive and the Waste Framework Directive, can be the catalyst that forces a change in behavior to direct MSW from the low cost disposal alternative.
- Even in the presence of a binding Directive, certain countries have ignored compliance.
- The recycle rate is not impeded by the presence of WTE whether it is a low or moderate value, further supporting the position that recycle rates are dependent on a national policy and not the presence of a WTE industry.

Waste Management for Sustainable Society. Table 3 Comparison of primary input and outputs of WARM and MSW-DST

Parameter	Units	MSW-DST	WARM
Input options			
Energy recovery	kWh/MMBtu input	Variable	Constant
Efficiency of landfill gas collection	%	Variable	Constant
Oxidation of methane by soil	%	Variable	Constant
Carbon grit intensity	kg CO ₂ /kWh	Variable	Constant
Carbon storage	kg C/kg MSW	Variable	Constant
Output values			
Energy saved/consumed	MMBtu	Yes	Yes
Greenhouse gas emissions	Tons	Yes	Yes
SO ₂ , NO _x , particulate	Tons	Yes	No
Solid waste	Tons	Yes	No
Water discharges	Tons	Yes	No

Life-Cycle Assessment (LCA) Methodology

The European Union formulates environmental directives according to a thematic strategy that is a multidisciplinary plan to assure that directives consider the big picture and interrelationship between different media including air, water, waste, toxics, etc. Their Integrated Pollution Prevention and Control [13] is an example of the multidisciplinary approach to decision making and introduces the concept of a life-cycle assessment (LCA) into the decision making process. A LCA is a long-term multidisciplinary evaluation methodology to assure that long-term implications are considered when selecting an option.

LCA methodology for waste management is endorsed by a variety of groups (USEPA, IPCC, Ceres) with there being a wide variety of LCA computational programs available in the market. The USEPA has two LCA's available for public use, the Waste Reduction Model (WARM) [14] and the Municipal Solid Waste Decision Support Tool (MSW-DST) [15]. EPA is currently reviewing each of these with a goal of creating consistency between the two. Table 3 provides a general comparison of the two and where they are different.

Evaluation of an Integrated Waste Management System

The following sections address several of the major parameters that warrant consideration when evaluating and/or designing an integrated waste management system.

- Energy – does the proposed system provide energy recovery, energy savings, or is it a sink?
- Greenhouse gases – does the proposed system provide mitigation (reduction) of greenhouse gases (GHGs) or is it a source?
- Land Management – can the proposed system be located in close proximity to the point of waste generation and what are the long-term impacts?
- Air emissions – does the system and its components provide for mitigation of air emissions including air toxics?

Energy is addressed first because it is a key parameter in determining the GHG and air emission characteristics.

Energy

A two-step approach is used to present and evaluate the energy parameter. Step 1 identifies the energy

savings/sink for specific components managed by various waste management options. Step 2 applies these component-based factors to national inventories as a means to present the energy recovery/savings for an integrated waste management system. A similar sequence is also followed with greenhouse gas (GHG) emissions in order to understand the relationship between energy and GHG emissions.

Derivation of Energy Factors

Primary energy represents all of the energy consumed, including the energy required to produce, generate, distribute, and transmit the electricity, petroleum, and natural gas used. This metric is used by the International Energy Association (IEA) and the US Energy Information Administration (EIA) and is well suited to compare different waste management options which generate electricity (WTE & LFGTE) and save electricity and fuels (recycling).

The primary energy savings or consumption for the three options was determined by USEPA for various waste components and composites through a life-cycle assessment (LCA). Table 4 presents USEPA's LCA energy factors [16] for various components managed by recycling, EfW, and landfills, with landfills being characterized by three types: vents (no gas collection and no energy recovery), flare (gas collection but no energy recovery), and energy (gas collection and energy recovery).

These energy factors are approximately 15–20 years old at this time and represent assumptions and operating conditions that may not necessarily represent existing conditions [16]. Several of the key parameters that should be considered when applied to a current situation are identified below according to the waste management option.

Recycling

1. There is one energy factor for a certain commodity regardless of its actual physical condition. As an example, there may be little variability in the energy savings when recycling aluminum cans, however, there may be significant differences when considering the variable quality of separated cardboard and paper, whether by contamination or being mixed with other materials. As a result – there is only one energy factor whether a country is separating 20% or 65% of a commodity.
2. Processes used to transform a separated commodity are likely to have changed in the last 20 years, however, any usullout danger the net energy difference between deriving a product from used and recovered items versus raw natural resources are not considered.
3. A large fraction of materials separated for “recycling” in North America and Europe are actually being “recycled” in China. The impact of energy savings in China versus other manufacturing locations is not considered by the USEPA factors.

Waste Management for Sustainable Society. Table 4 Energy factors for discrete components and mixed recyclables (mmbtu/t)^a

Component	Recycling	WTE	LF-vent	LF-Flare	LF-Energy
Aluminum cans	−206.42	0.65	0.53	0.53	0.53
Steel cans	−19.97	−17.09	0.53	0.53	0.53
HDPE	−50.90	−19.61	0.53	0.53	0.53
Cardboard	−15.05	−7.21	0.53	0.53	−0.56
Office paper	−10.08	−6.95	0.53	0.53	−1.31
Mixed recyclables	−16.57	−7.22	0.53	0.53	−0.27
Compost	0.58	−2.32	0.53	0.53	−0.01
Recyclables and compost	−11.92	−5.89	0.53	0.53	−0.20

^aSource WARM 10 using default values [14]

Waste-to-Energy (WTE; or Energy from Waste, EfW)

USEPA's factors are based on operating characteristics of WTE facilities 20 years ago when the evaluation was completed. As a result, the estimated energy generation from WTE in WARM is low for several reasons:

1. Net energy distributed to the grid is based on a nominal thermal energy efficiency of 18% which is quite low in comparison to new units in Europe and North America that operate at up to 30% thermal efficiency (electricity generation only) and 90% when they also provide steam for local district heating.
2. Nearly all existing WTE facilities and *all* new WTE facilities recover both ferrous and nonferrous metals. This increases the energy savings because these materials are used in downstream recycling operations.

Landfills USEPA has a separate factor for three different landfill designs: landfill gas collection with vents, collection and flaring, and collection with combustion in a reciprocating engine. There are no expected changes to energy savings for landfills with vents or flares, however, a landfill that recovers more gas should yield more energy per ton of MSW buried. The absence of verifiable and proven long-term collection efficiencies and power generation make any improvements in energy generation speculative.

The results in [Table 4](#) follow a logical pattern when evaluating component specific factors. As an example, it is common knowledge that recycling aluminum saves large amounts of energy but aluminum is a small weight percent of MSW and does not have much of an influence on a standard basket of mixed recyclables where the results are driven by large amounts of paper, plastic, and cardboard. A different and independent analysis by the Earth and Environmental Engineering Department of Columbia University [17] yielded similar results using a Resource Conservation Efficiency (RCE) metric that also placed energy savings of recycling ahead of WTE with landfills being in last place.

The energy factor associated with recycling represents the energy impact of collecting, processing, and remanufacturing the component. The "mixed recyclables" factor is based on USEPA's weight distribution which is limited to conventional materials including

paper, glass, metals, and plastic. Compostable materials are treated as a separate line item and then combined with mixed recyclables.

Several major observations regarding [Table 4](#) are:

- According to this data base, the ratio of energy saved by recycling to EfW is 2.2 for a mixed basket of recyclables ($-16.57/-7.22$) and 2.0 for one including both recyclables and compostable materials ($11.92/5.89$). This ratio is based on the assumptions identified above.
- The amount of energy saved by recycling is about 60 times greater than that recovered by a landfill with energy recovery whether the waste basket is recyclables ($16.57/0.27$) or recyclables and compost ($11.92/0.2$). Landfill with vents and flares are energy sinks (consume energy). This makes sense when considering the energy required for transportation and landfill management over the nominal 100 year anaerobic decomposition period for MSW.
- The amount of energy recovered/saved by EfW is 27–30 times more than landfills with energy recovery when considering recyclables or recyclables and compost.

The lifecycle assessment factors in [Table 4](#) from USEPA's WARM model the different time requirements for each option. Both recycling and EfW are relatively short-term processes with EfW being a nominal 1-h process after MSW is charged to the grate. Recycling is also a short-term process when considering the time to transfer separated components to a transformation process where those components are converted to a recycled commodity. Landfills are very long-term processes due to the nominal 100 year anaerobic decomposition process. Landfills with gas collection and energy recovery generate electricity for export for a certain period when gas collection is of certain quantity and quality (25–35 years), however, there are energy demands for the entire life of the landfill. Independent analysis estimate that an average existing EfW facility generates 10 times the electricity delivered to the grid than a landfill gas to energy system with new EfW facilities using combined heat and power being significantly greater. As noted above, USEPA's LCA energy analysis demonstrates that a typical EfW is 27–30 times more effective than a landfill gas to energy option.

Application of Energy Factors

The approach used herein to assess the primary energy savings from an integrated waste management system is to compare results from two scenarios: Business as Usual (BAU) which is the situation today in each geographic area and a Sustainability Scenario (SUS) where an integrated waste management system is used with waste allocated according to the hierarchy. This approach allows a direct estimation of the energy benefits today if society simply changed the way it manages MSW. Table 5 presents the total amount of MSW (as millions of tons) and the percentage managed by each option. The BAU values are derived from a variety of sources whereas the distribution of MSW in the SUS scenario is based largely on the current situation in Germany where they have managed to achieve a nominal 65% recycle rate. The recycle rate in Germany (which is based on separating conventional materials (paper, plastic, metals, glass, cardboard, and composting) is the best in the EU.

The three examples (EU, USA, and East Asia (China)) were selected for several reasons:

- The EU provides the longest regulatory effort to advance recycling on a large scale. However despite several countries leading the way, the average value demonstrates that there is still room for improvement albeit less than the other two areas.
- The population and total waste of North America and the EU is about the same but the management of MSW is very different with the USA having a relatively stagnant waste management system compared to the European Union where diversion of MSW from landfills is considered to be a sustainable solution.

- China is the world's largest generator of MSW and has already demonstrated that its market, or absence thereof, for separated components can influence reported recycle rates. China, like the European Union, has also determined that landfills are the least sustainable option and is moving toward a sustainable integrated waste management system, however, they have not established a goal similar to Germany's operating condition, therefore there continues to be significant room for improvement.

Table 6 presents a comparison of the energy savings of the BAU and SUS Scenario using MSW rates listed in Table 5.

In order to provide some perspective on the magnitude of Quads, the Department of Energy [18] estimated the USA's 2008 energy use as 99.2 Quads with Table 7 providing a breakdown of that usage and the amount allocated to generating electricity.

Several observations regarding Tables 5, 6, and 7:

- Despite landfills being an energy sink – the majority of MSW in the EU, USA, and China is landfilled.
- Diversion of MSW from landfills to recycling and EfW provides a significant energy savings through the use of existing technology.
- The total energy savings of 2.2 Quads in the USA from the SUS Scenario is approximately 2.2% of the total energy use in the USA and 5.6% of that allocated to generating electricity. Approximately 2.2 Quads is the amount from hydro facilities and could offset a large fraction of electricity from coal and natural gas generation.

The amount of potential energy that can be saved/recovered from a sustainable allocation of MSW is

Waste Management for Sustainable Society. Table 5 Allocation of MSW for business-as-usual (BAU) and sustainability scenario (SUS)

Reference information		BAU scenario			SUS scenario		
Region	2004 ^a	Landfill	EfW	Recycle	Landfill	EfW	Recycle
Europe	379	58	12	26	12	23	65
North America	371	68	6	26	12	23	65
East Asia	442	78	12	10	20	40	40

^aMSW values (as million tons) and allocation as % of total is from [6]

Waste Management for Sustainable Society. Table 6 Primary energy savings (as Quads) for the business-as-usual (BAU) and sustainable scenario (SUS)^a

MSW option	EU			USA			China		
	BAU	SUS	Delta	BAU	SUS	Delta	BAU	SUS	Delta
Landfill	0.0	0.0	0	−0.1	0	0.1	−0.1	0.0	0.1
Recycle	1.2	2.8	1.5	1.1	3.0	1.8	0.5	3.6	3.0
WTE	0.3	0.5	0.3	0.1	0.6	0.5	0.2	0.3	0.2
Total	1.5	3.3	1.8	1.1	3.5	2.4	0.7	3.9	3.3

^a1 Quad = 10¹⁵ Btu

Waste Management for Sustainable Society. Table 7
Estimated US energy use in 2008 (as Quads)

Source	Total	Electricity generation
Solar	0.09	0.01
Nuclear	8.45	8.45
Hydro	2.45	2.43
Wind	0.51	0.51
Geothermal	0.35	0.31
Natural gas	23.84	6.82
Coal	22.42	20.54
Biomass	3.88	0
Petroleum	37.13	0
Total	99.12	39.07

significant on an absolute and relative basis. For the USA alone, the energy savings is equivalent to over 10% of our 2008 total oil imports or equivalent to the coal consumption of 63,500 MW coal plants. For the USA, Europe, and China together, the energy savings is equivalent to the coal consumption of nearly 200 large 500 MW coal plants.

An independent analysis by Columbia University Scientists evaluated the energy savings with their own metric called the Resource Conservation Efficiency (RCE) [22]. The RCE metric was applied to two leading cities in the USA, San Francisco and Honolulu, to evaluate their sustainable use of MSW as a resource according to their RCE metric. San Francisco has a higher recycle rate (34.4%) than Honolulu (27.2%) however 65.6% of MSW was landfilled yielding an RCE of 60.4. Honolulu minimized landfill (4.9%) with WTE

managing 67.9% for a total RCE of 73.1. While Columbia's RCE metric is a different energy-based approach to evaluate sustainable use of MSW as a resource – its conclusions are consistent on a component basis and when applied to a larger source.

Subsequent sections consider the environmental impacts when diverting MSW from landfills to recycling and EfW.

Greenhouse Gas Emissions

Overview

Climate change has been correlated to an increase in the concentration of greenhouse gases (GHGs) driven predominantly by anthropogenic emission sources. The most comprehensive report on the topic is the Fourth Assessment Report by the Intergovernmental Panel on Climate Change (IPCC) that included several key findings applicable to the topic of sustainable waste management:

- Increases in atmospheric CH₄ concentrations since preindustrial times have contributed a radiative forcing of $+0.48 \pm 0.05 \text{ W m}^{-2}$. Among greenhouse gases, this forcing remains second only to that of CO₂ in magnitude.
- Emissions of CO₂ from fossil fuel use and from the effects of land use change on plant and soil carbon are the primary sources of increased atmospheric CO₂.
- Current understanding of anthropogenic warming and cooling influences on climate has led to a very high confidence (at least 9 out of 10 chance) that the effect of human activities since 1750 has been a net positive radiative forcing of $+1.6 (+0.6 \text{ to } +2.4) \text{ W m}^{-2}$.

- It is very likely that anthropogenic greenhouse gas increases caused most of the observed increase in global average temperatures since the mid-twentieth century.

It is readily apparent from these IPCC findings that processes that reduce GHG emissions, including emissions of CO₂ and CH₄, should be preferred relative to attempts to capture and sequester GHG emissions after they have been generated. In other words, prevention of GHG emissions at the source is more effective than capturing and controlling GHG emissions.

Atmospheric measurements have demonstrated that the concentration of atmospheric CO₂ has increased from a preindustrial value of about 280 to 379 ppm in 2005 with future concentrations being dependent on a variety of complicated factors, one key factor being human behavior and the source of electrical energy to meet growing demand. In 2004, the journal *Science* published a pivotal paper by Pacala and Socolow that simplified the situation by considering two situations over a 50 year period, one where emissions are flat and one where emission increase according to the existing rate. This yields a right triangle or “stabilization wedge” with a growth of 7 billion tons of carbon over a 50 year period. The right triangle could then be broken down into more manageable “wedges,” each of which could achieve a reduction in GHG emissions of at least 1 gigaton of carbon per year

(GtC/year) by 2054 [19]). Seven 1 t wedges would stabilize carbon emissions. Fifteen stabilization wedges were presented in six categories: energy efficiency (vehicles, buildings, power plants), fuel shift (switching from coal power to nuclear, wind or natural gas), CO₂ capture and storage, nuclear fission, renewable energy (wind, photovoltaic, biogas), and forest and agricultural soils (reduced deforestation and tillage).

Individual waste management options and an integrated system were not considered by this initial effort. However, application of an integrated solid waste management policy based on the waste hierarchy was subsequently identified as a stabilization wedge [6].

GHG Emission Characteristics of Waste Management Options

Each of the three waste management options considered in the energy section has GHG emission characteristics that must be considered in any discussion on sustainability. Table 8 presents an overview of the general GHG characteristics of each relative to CO₂ and CH₄ emissions.

Table 9, which presents GHG emission characteristics for specific waste components, is organized to be consistent with Table 4 which presents the energy factors for each component.

The GHG emission characteristics in Table 9 are absolute values for each option using a lifecycle

Waste Management for Sustainable Society. Table 8 Overview of GHG emission characteristics

Option	GHG	Characteristic	Mechanism
Recycling and composting	CO ₂	Mitigation	Energy savings = avoided fossil CO ₂ ^b
	CH ₄	Mitigation	Avoided landfill CH ₄
Energy from waste	CO ₂	Source	Combustion of plastics cutler anthropogenic also
	CO ₂	Mitigation	Recovery of metals for recycling saver energy
	CO ₂	Mitigation	Power generation = avoided fossil CO ₂ ^{a,b}
	CH ₄	Mitigation	Avoided landfill CH ₄
Landfill (disposal)	CO ₂	Mitigation or source	Mitigation for landfills with energy recovery – avoided fossil CO ₂
	CH ₄	Source	All landfills emit CH ₄ with the amount dependent on the effectiveness of the landfill gas collection system

^aDoes not include energy savings from recovery of metals

^bDoes not include upstream emissions from fuel production

Waste Management for Sustainable Society. Table 9 GHG emissions as Mg CO₂e/Mg MSW

Material	Recycling			EfW ^a	Landfill ^b
	Energy	Carbon storage	Total		
Aluminum cans	−0.37	0.0	−3.7	+0.02	+0.01
Steel cans	−0.49	0.0	−0.49	−0.42	+0.01
HDPE	−0.38	0.0	−0.38	+0.28	+0.01
Cardboard	−0.02	−0.83	−0.85	−0.17	+0.11
Office paper	0.05	−0.83	−0.78	−0.16	+0.53
Mixed recyclables	−0.13	−0.66	−0.79	−0.15	+0.05
Mixed MSW	Not applicable			−0.04	+0.41

^aFactor assumes EfW does not include nonferrous recovery

^bWARM version 10 using default values national averages, CH₄ emissions based on 75% collection and C sequestration

assessment (EPA WARM). Any one megagram (metric ton) of MSW or a discrete component can only be managed by one option, therefore difference between LCA results for two options is the net advantage or disadvantage relative to climate change. Observations from Table 9 on discrete components and different options are:

- Recovery of both ferrous and nonferrous metals is an effective GHG mitigation strategy.
- Management of paper and cardboard by recycling and EfW are GHG mitigation strategies, however, the recycling factors represent the high-end of separated commodities. The energy savings may not be the same for lower-end or contaminated materials.
- Landfills are a net source of GHG emissions even when assigning a favorable lifetime landfill gas collection efficiency of 75%.

The two GHG components of recycling, energy and carbon sequestration, are presented to highlight the value of each mechanism. Recycling is generally recognized as a demonstrated GHG mitigation technology because it can reduce GHG emissions at the manufacturing stage (saves energy), increase forest carbon sequestration, and avoid landfill CH₄ emissions, however close inspection shows that, depending on the material recycled, the majority of GHG mitigation can be from forest carbon sequestration.

While it is clear that recycling is a preferred management option for both energy savings and GHG

mitigation – what is the best option for MSW that is not, or cannot be, recycled? Information in Tables 4, 5, and 6 demonstrated that WTE is the superior alternative when considering energy. Table 10 presents GHG emission factors on a lifecycle assessment basis when using USEPA's MSW-DST to compare the impact of managing one Mg of MSW by either EfW or the nationwide blend of landfills in the USA. The MSW-DST was used in this analysis because it allows the user to vary key parameters (methane generation rate, landfill gas collection efficiency, etc.) whereas WARM does not.

Observations from Tables 9 and 10 are:

- Recycling is a demonstrated GHG mitigation technology based on three different processes (save energy, forest carbon sequestration, and avoid landfill methane) with the exact values depending on local conditions.
- EfW is a demonstrated GHG mitigation technology based on energy recovery and avoided landfill methane with the exact values also depending on local conditions.
- Landfills are a source of GHG emissions with the exact amount being highly variable and dependent on sustained long-term performance at the landfill.
- A comparison of EfW and various landfill options demonstrates that EfW reduces GHG emissions with the amount being dependent on local conditions.

Waste Management for Sustainable Society. Table 10 Nationwide life cycle greenhouse gas emission factors for MWC versus landfill blend scenarios^a

Integrated (lifetime) landfill gas collection efficiency	GHG emissions (Mg CO ₂ e/Mg MSW)			
	Landfill allocation (50/25/25) ^a		Landfill allocation (40/30/30) ^b	
	Lo = 100	Lo = 170	Lo = 100	Lo = 170
25	1.44	2.20	1.4	2.14
35	1.37	2.08	1.32	2.00
45	1.30	1.96	1.23	1.85
55	1.23	1.84	1.15	1.71
65	1.16	1.72	1.06	1.56
75	1.08	1.66	0.98	1.42

^a50% vent, 25% flare, 25% with energy recovery

^b40% vent, 30% flare, 30% with energy recovery

Application of GHG Emission Factors

Application of an integrated solid waste management policy based on the waste hierarchy has been identified as a stabilization wedge [6]. The analysis demonstrated that if MSW is allocated to recycling, waste-to-energy, and landfilling in descending order in lieu of existing “business-as-usual” practices with each option using modern technology and best practices, the system would reduce greenhouse gas emissions by more than 1 GtC/year. This integrated waste management system reduces CO₂ by displacing fossil electrical generation and avoiding manufacturing energy consumption and methane emissions from landfills.

As mentioned above, the concept of a stabilization wedge is to examine the GHG mitigation over a 50 year period on the premise that it takes that much time to implement change. Table 11 presents the GHG mitigation associated with immediate implementation of an integrated waste management system using MSW data in Table 4. These GHG mitigation results represent energy savings described in Table 6.

GHG mitigation is accomplished by two factors: (1) diverting MSW from landfills to recycling and EfW and (2) use of technology consistent with existing best practices. Diversion of MSW from landfills avoids methane emissions and by allocating MSW to recycling and EfW, there are significant energy savings and therefore avoidance of fossil CO₂ emissions.

Waste Management for Sustainable Society. Table 11 GHG mitigation attributable to an integrated waste management system

Land area	GHG mitigation	
	Million metric tons of C	Million metric tons as CO ₂ e
Europe ^a	40	148
North America ^a	60	220
East Asia ^a	94	337
Global ^b	1,009	

^aBased on 2004 MSW tonnage and management

^bBased on 2054 MSW tonnage and management

The original wedge analysis was a global assessment that considered a baseline condition and several sensitivity calculations. Table 12 provides an overview of the results.

The GHG mitigation for Europe, North America, and East Asia in the Waste Wedge was 0.41 GtC or about 41% if the total mitigation when considering growth of waste quantities by 2054. An immediate reduction of existing MSW in the three geographic areas, based on 2004 data, indicates an immediate

Waste Management for Sustainable Society. Table 12
Overview of integrated waste management as a stabilization wedge

Item	Reference	Results as GtC
Waste wedge	Baseline using 100-year GWP of 21	1.009
A	After maximizing recycling, maximize EfW	1.094
B	After maximizing recycling, minimize EfW	0.763
C	Maximize thermal efficiency of EfW in Waste Wedge	1.022
D25	Apply 100-year GWP of 25 for methane instead of 21	1.153
D72	Apply 20-year GWP of 72 for methane instead of 21	2.842
E	Include carbon storage factor in recycling	2.132
F	Add source reduction in waste wedge analysis	1.251

reduction of about 0.2 GtC. This is a conservative and significant factor. It is conservative because:

- It is using a 100-year Global Warming Potential of 21 whereas the IPCC has raised the 100-year GWP to 25 and others recommend a 100-year GWP of 34 [20].
- The 20-year GWP of 72 is more appropriate when considering near term actions which can have an immediate impact on radiative forcing.

It is significant when considering the equivalencies. According to USEPA's equivalency calculator, 0.4 GtC is equivalent to the annual GHG emissions:

- Of approximately 350 large coal fired power plants,
- Of over 285 million automobiles, or
- From the energy consumption of every home in the USA.

Emissions

A direct comparison of emission from the three options is difficult for a variety of reasons. First, recycling of various separated commodities occurs at different operations (pulp and papermills, aluminum smelters, arc furnaces,

Waste Management for Sustainable Society. Table 13
Overview of monitoring and testing requirements of WTE

Pollutant	European Union WID	USEPA CAAA
CEM		
Sulfur dioxide	Yes	Yes
Nitrogen oxides	Yes	Yes
Carbon monoxide	Yes	Yes
Particulate	Yes	Yes
Hydrogen chloride	Yes	^a
Hydrogen fluoride	Yes	^a
Hydrocarbons	Yes	^a
Manual testing		
Metals	Yes	Yes
Dioxins and furans	Yes	Yes

^aNot a national federal requirement however facility specific permits may include this requirement

etc.), each of which has a different set of standards in the USA, and with much material being exported to China, emissions are especially difficult to estimate. For MSW that is not recycled – regulations in the EU, USA, and elsewhere have led to a robust data set useful for characterizing WTE emissions. In contrast, landfills have not been subject to a comprehensive set of standards that requires testing and reporting of characteristic emissions.

WTE has been subject to extensive regulation in the EU through the Waste Incineration Directive (WID) [21] and through the Clean Air Act Amendments of 1990 (CAAA) in the USA [22]. While each set of regulatory standards has undergone a distinct history, the results are similar in that each set of regulations has identified the same proximate set of pollutants as being appropriate for effective management if emissions after 100% of flue gas have been managed by modern air pollution control technology. Both the WID and CAAA require continuous emissions monitoring (CEM) and periodic stack testing of other pollutants where CEM is not technically feasible with Table 13 presenting an overview of these requirements.

A directly comparable national standard for WTE was not available for China and Canada, however, WTE

Waste Management for Sustainable Society. Table 14 Landfill testing requirements in the European Union and the USA

Parameter	European Union landfill directive [23]	USEPA [22]
Gas collection	All landfills that accept biodegradable waste	Only those with NMOC emission above 50 mg per year
Gas collection period	Implies forever	Shutdown allowed 15 years after initial startup of system
Testing – bulk emissions (CH ₄ , CO ₂ , O ₂ , N ₂ , H ₂ S)	Operating landfill – monthly	None
	Closed landfill – Semiannual	
Testing – trace emissions		
Air toxics	No	No
Known carcinogens	No	No
Probable carcinogens	No	No
CEM required	No	No

projects in China are being permitted in accordance with the EU WID and recent projects in Canada have also been permitted in accordance with the EU WID or more stringent standards.

Continuous or intermittent measurement of emission from landfills is relatively complicated due to there being two basic sources: (1) an area source of landfill gas that is not collected by the landfill gas collection system and (2) emissions from the flare and/or engine where collected gas is combusted. Neither EU nor USEPA regulations require periodic testing of fugitive emission from landfills. Only minimal testing or monitoring (largely flue gas temperature) of combustion gas from the flare or engine is required. EU regulations are more rigorous, apparently due to the overall objective being to reduce negative impacts to the local environment (surface water, groundwater, soil and air) and global environment including the greenhouse effect and any resulting risk to human health during the whole life-cycle of the landfill. However, the EU standards do not require routine stack testing or monitoring in a manner similar to WTE.

Table 14 provides an overview of landfill test requirements in EU and USEPA.

USEPA's draft update of emission factors for landfills [24] listed 170 air pollutants of which 45 are Title III air toxics including several known carcinogens and many more reasonably anticipated carcinogens.

A comparison of landfill emissions with either recycling or WTE operations is difficult because landfills are not required to test, resulting in a minimal data base.

A third source of emissions from landfills that warrants mention is surface and underground landfill fires. According to USEPA [25] and the Federal Emergency Management Agency (FEMA), [26] both types of landfill fires are known to be the source of emissions including dioxins. FEMA has estimated 8,400 landfill fires per year based on data from the National Fire Incident Reporting System (NFIRS) with USEPA estimating a possible dioxin emission of 1,100 g of toxic equivalence per year. At this estimated level, landfills would be the largest anthropogenic source of dioxin emission in the USA when compared to a recent USEPA dioxin inventory [27].

Land Management

Each of the three MSW management options has a land impact, however in a manner similar to emission, there is little data available to quantify the difference between each.

USEPA's life-cycle assessment of MSW management options recognized that when any material is separated for recycling, some portion of that material is unsuitable for use and is discarded. The loss rates [14] for different separated commodities were reported with

a range of 0.34 (textbooks) and 0.01 (copper wire). The typical mixed basket of recyclables, excluding compost, yields a loss factor of about 0.15, meaning that for every ton of recycled products there is 0.15 ton of material that must be managed by either WTE or landfill.

The automatically controlled combustion process at a modern WTE facility converts a ton of MSW into 0.25 ton of total inert residue in a nominal 1-hour period. This reduction in mass is proportional to a 90% reduction in volume. This essentially extends the lifetime of a landfill by a factor of 10 if both a WTE and landfill were to accept the same amount of material for any given period of time.

Summary

There is extensive information that documents that an integrated waste management system that manages municipal solid waste as a resource would be a sustainable solution for both generation of renewable energy and reduction of greenhouse gas emissions. Commercially available technology is available, however there are not sufficient existing policies and regulations to change dependence on landfills, the low cost option that has the least sustainable attributes.

Diversion of MSW from landfills to recycling and WTE systems would provide significant energy savings and a significant reduction in the emissions of GHG's. The European Union's establishment of directives is an example of how this can be accomplished.

Bibliography

- European Parliament; Council of the European Union (2008) Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain Directives. *J Eur Union L* 312/3
- U.S. EPA (2009) Opportunities to reduce greenhouse gas emissions through materials and land management practices. U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response
- Kaufman SM, Themelis NJ (2009) Using a direct method to characterize and measure flows of municipal solid waste in the United States. *J Air Waste Manag Assoc* 59:1386–1390
- United Nations World Commission on Environment and Development (WCED), 1987, *Our Common Future*, United Nations, New York. <http://www.ceres.org/about-us/our-history>
- European Union (EU) (2002) Decision No 1600/2002/EC of the European Parliament and of the Council of 22 July 2002 laying down the Sixth Community Environment Action Programme. *Off J Eur Communities L242*:1–15
- Bahor B, Van Brunt M, Blue K, Stovall J (2009) Integrated waste management as a climate stabilization wedge. *Waste Manag Res* 27:839–849
- European Union (EU) Eurostat, Municipal waste generated and treated. http://epp.eurostat.ec.europa.eu/portal/page/portal/waste/data/sectors/municipal_waste. Accessed 25 Sept 2011
- USEPA (2010) Municipal solid waste in the United States: 2009 facts and figures. EPA530-R-10-012. United States Environmental Protection Agency Office of Solid Waste (5306P), Washington, DC
- Arsova L, Van Haaren R, Goldstein N, Kaufman S, Themelis N (2008) 16th Nationwide Survey of MSW Management in the US: The State of Garbage in America. *BioCycle*
- Van Haaren R, Themelis N, Goldstein N (2010) 17th Nationwide Survey of MSW Management in the US: The State of Garbage in America. *BioCycle*
- USEPA (2011) Inventory of U.S. GHG Emissions and Sinks: 1990–2009. USEPA
- Berenyi E (2009) Recycling and waste to energy: are they compatible? 2009 update
- European Commission (2006) Integrated pollution prevention and control – reference document on the best available technologies for waste incineration
- US EPA (US Environmental Protection Agency) (2006) Waste Reduction Model (WARM). Version 10, November 2009. http://epa.gov/climatechange/wycd/waste/calculators/downloads/WARM_v10.zip. Accessed 17 June 2011
- US EPA (US Environmental Protection Agency) (2002) Fact sheet: cutting edge software to cut emissions. <http://www.epa.gov/osw/nonhaz/municipal/pubs/ghg/f02024.pdf>. Accessed 25 Sept 2011
- US EPA (2006) Solid waste management and greenhouse gases – a lifecycle assessment of emissions and sinks, 3rd edn. U.S Environmental Protection Agency, Washington, DC
- Kaufman SM, Krishnan N, Themelis N (2010) A screening lifecycle metric to benchmark the sustainability of waste management systems. *Environ Sci Technol* 44:5949–5955
- Lawrence Livermore National Laboratory (2008) U.S. Energy flowchart 2008. <https://www.llnl.gov/news/newsreleases/2009/images/energy-use.pdf>. Accessed 25 Sept 2011
- Pacala S, Socolow R (2004) Stabilization wedges: solving the climate problem for the next 50 years with current technologies. *Science* 305:968–972
- Shindell DT, Faluvegi G, Koch DM, Schmidt GA, Unger M, Bauer SE (2009) Improved attribution of climate forcing to emissions. *Science* 326:716–718
- European Union (EU) (2000) Directive 2000/76/EC of the European Parliament and of the Council of 4 December 2000 on the incineration of waste. *Off J Eur Communities L332*: 91–111
- US EPA (Environmental Protection Agency) (1996) Standards of performance for Municipal Solid Waste Landfills. *Fed Regist* 49(61):9905–9944

23. Council Directive 1999/31/EC of 26 April 1999 on the landfill of waste. Official Journal of the European Committee, 16 Jul 1999
24. U.S. EPA (Environmental Protection Agency) (2008) Draft update, compilation of air pollutant emission factors, Section 2.4, Municipal Solid Waste Landfills, Oct 2008
25. U.S. EPA (Environmental Protection Agency) (2006) An inventory of sources and environmental releases of dioxin-like compounds in the United States for the years 1987, 1995, and 2000. National Center for Environmental Assessment, Washington, DC, EPA/600/P-03/002 F. National Technical Information Service, Springfield. <http://epa.gov/ncea>
26. FEMA (Federal Emergency Management Agency) (2002) Landfill fires: their magnitude, characteristics and mitigation. Federal Emergency Management Agency United States Fire Administration National Fire Data Center, May 2002/FA-225
27. US EPA (2006) An inventory of sources and environmental releases of dioxin-like compounds in the United States for the years 1987, 1995, and 2000. US EPA, Washington, DC

Waste Materials in Construction, Utilization of

MAGDY ABDELRAHMAN, ARASH SABOORI

Department of Civil Engineering (CIE 201F), North Dakota State University, North – Fargo, ND, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Recycling

Recycling As a Rehabilitation Alternative

Types of Recycled Materials and Related Statistics

Applications of Waste Materials in Building

Construction

Applicability of Current Specifications to Recycled Materials

Future Directions

Acknowledgments

Bibliography

Glossary

By-product A secondary or incidental product deriving from a manufacturing process, a chemical reaction, or a biochemical pathway, and is not the primary product or service being produced.

Construction and Demolition Debris A large and varied waste stream that includes concrete, asphalt, wood, gypsum, and asphalt shingles generated from the construction, renovation, and demolition of buildings, roads, bridges, and dams.

Industrial Materials By-products of industrial processes.

Recycling Processing used materials into new products to prevent waste of potentially useful materials, reduce the consumption of fresh raw materials, reduce energy usage, reduce air pollution and water pollution by reducing the need for “conventional” waste disposal, and lower greenhouse gas emissions as compared to virgin production.

Rehabilitation Measures to improve, strengthen, or salvage existing deficient pavements to continue service with only routine maintenance. Deficient pavements exhibit distress in excess of what can be handled through routine maintenance.

Sustainability The capacity to endure. For humans, it is the potential for long-term maintenance of well-being, which in turn depends on the well-being of the natural world and the responsible use of natural resources.

Sustainable development A pattern of resource use that aims to meet human needs while preserving the environment so that these needs can be met not only in the present, but also for future generations.

Definition of the Subject

Recycling has always been a matter of obtaining an economic benefit or of helping the environment. Until the Industrial Revolution when materials became more available than labor, reuse and recycling was commonplace. Nearly 4,000 years ago, there was a recovery and reuse system of bronze scrap in operation in Europe and there is evidence that composting was carried out in China. Since the industrial revolution, the waste production rate has grown greatly. The idea of landfills came about in the 1940s and 1950s. When these huge depositories became available, they were extremely popular because of the ease of tossing unused products away. No one anticipated that soon these landfills would multiply to the extent that they have.

Considering the enormous volume of construction work done annually and the huge amount of waste materials produced each year, the issue of reusing the waste materials in construction is gaining increased attention, which can be observed by the extensive research being done on the subject. These wasted materials are either debris and demolition materials from construction industry or waste produced by other industries. The rapid growth of industries results in the need for more raw materials and the production of more waste materials, which makes the use of waste materials more feasible from two aspects, environmental and economical.

Introduction

Recycling has been a common practice throughout human history. In preindustrial times, scrap made of bronze and other precious metals was collected in Europe and melted down for perpetual reuse, and in Britain, dust and ash from wood and coal fires was downcycled as a base material in brick making. The main driver for these types of recycling was the economic advantage of obtaining recycled feedstock instead of acquiring virgin material.

Resource shortages caused by the world wars, and other such world-changing occurrences greatly encouraged recycling. Massive government promotion campaigns were carried out in World War II in every country involved in the war, urging citizens to donate metals and conserve fiber, as a matter of significant patriotic importance.

In the USA, the next big investment in recycling occurred in the 1970s, due to rising energy costs [1]. A main concern to deal with is the increasing volume of landfills due to large quantity of waste which are not recycled. Landfill usage peaked in the 1980s, when Americans sent almost 136 million metric tons (150 million US tons) of garbage to landfills each year. Today, Americans still dump more than 91 million metric tons (100 million US tons) of trash into landfills annually. However, recycling efforts in the United States divert 32% of waste away from landfills [2].

Among all the waste materials, two categories are of main concern in this entry, Industrial Waste and Construction and Demolition Materials (C&D) [3].

Construction and Demolition (C&D) waste accounts for a large part of the waste stream in the United States. While most of this waste accumulates in landfills, experts estimate that 90% of this waste stream is potentially reusable or recyclable. Recycling this waste can help to prolong our supply of natural resources and save money in the process [4]. Total US C&D waste was estimated to be 295 million metric tons (325 million US tons) in 2003 not including garbage.

Reducing and recycling C&D materials conserves landfill space, reduces the environmental impact of producing new materials, creates jobs, and can reduce overall building project expenses through avoided purchase/disposal costs [3]. Recycled Materials can be used in different categories of construction industry; their application in building and pavement as alternatives to virgin materials is discussed here.

Recycling

Literature defines recycling as processing used materials into new products to prevent waste of potentially useful materials, reduce the consumption of fresh raw materials, reduce energy usage, reduce air pollution (from incineration) and water pollution (from land filling) by reducing the need for “conventional” waste disposal, and lower greenhouse gas emissions as compared to virgin production [5].

Recycling or reuse of pavement materials is a powerful concept. Because of the reuse of existing material, pavement geometrics and thickness can also be maintained during construction. In some cases, traffic disruption is less than that for other rehabilitation techniques. Literature lists the specific benefits of recycling as:

1. Reduced costs of construction
2. Conservation of aggregate and binders
3. Preservation of the existing pavement geometrics
4. Preservation of the environment
5. Conservation of energy
6. Less user delay

Recycling as a Rehabilitation Alternative

Several studies have shown that it costs highway agencies less if the pavements are kept at a certain acceptable level of serviceability. According to World

Bank Sources, each \$1.00 spent during the first 40% drop in quality will cost \$4.00–\$5.00 if delayed until the pavement loses 80% of its original quality. Rehabilitation is needed to maintain the pavement at a certain condition. Rehabilitation of pavement is also needed due to the following reasons:

1. Inadequate ride quality
2. Excessive pavement distress
3. Reduced surface friction
4. Excessive maintenance requirement
5. Unacceptable user costs
6. Inadequate structural capacity for planned use or projected traffic

Recycling is only one of the several rehabilitation alternatives available for asphalt pavements. Some of the other common methods are thick or thin hot mix asphalt (HMA) overlay. The choice of rehabilitation alternative depends on observed pavement distress, laboratory and field evaluation of existing material, and design parameters. Also, maintenance of geometrics and original thickness of pavements, especially in underpasses, influence the choice of rehabilitation method. However, recycling has some unique advantages which are not available with other types of rehabilitation techniques. For example, recycling can result in savings, help in conservation of natural resources, and can maintain pavement geometrics as well as thickness. Different recycling methods are now available to address specific pavement distress and structural needs.

Recycling presents environmental opportunities and challenges, which, when appropriately addressed, can maximize the benefits of reuse. The use of most recycled materials poses no threat or danger to the air, soil, or water. Furthermore, careful design, engineering, and application of recycled materials can reduce or eliminate the need to search for and extract new, virgin materials from the land. The engineering feasibility of using recycled materials has been demonstrated in research, field studies, experimental projects, and long-term performance testing and analysis. Significant advances in technology over the past decade have increased the types of recycled materials in use and the range of their applications. When appropriately used, recycled materials can effectively and safely reduce cost, save time, offer equal or, in some cases,

significant improvement to performance qualities, and provide long-term environmental benefits [6].

The Federal Highway Administration (FHWA) policy toward recycling is [7]:

1. Recycling and reuse can offer engineering, economic, and environmental benefits.
2. Recycled materials should get first consideration in materials selection.
3. Determination of the use of recycled materials should include an initial review of engineering and environmental suitability.
4. An assessment of economic benefits should follow in the selection process.
5. Restrictions that prohibit the use of recycled materials without technical basis should be removed from specifications.

Types of Recycled Materials and Related Statistics

The different recycled and by-product materials which have application in construction industry are discussed and an estimation of annual production of each is presented [8–10].

Baghouse Fines

Hot mix asphalt baghouse fines are dust particles that are captured from the exhaust gases of asphalt-mixing plants. Secondary collection equipment called baghouses is commonly used to capture these very fine-sized materials. There are approximately 3,600 hot mix asphalt plants in the United States, accounting for a total annual production of 400–450 million metric tons (450–600 million US tons) of asphalt paving material. About 2,300 of these hot mix plants are batch plants, with the remaining 1,300 being drum mix plants. Roughly 40–50% of all hot mix asphalt plants are equipped with baghouse collection systems. It is estimated that approximately 5.4–7.2 million metric tons (6–8 million US tons) of baghouse fines are generated annually by the US asphalt production industry.

Blast Furnace Slag

In the production of iron, iron ore, iron scrap, and fluxes (limestone and/or dolomite) are charged into

a blast furnace along with coke for fuel. The coke is combusted to produce carbon monoxide, which reduces the iron ore to a molten iron product. This molten iron product can be cast into iron products, but is most often used as a feedstock for steel production. Blast furnace slag is a nonmetallic coproduct produced in the process. It consists primarily of silicates, aluminosilicates, and calcium-alumina-silicates. The molten slag, which absorbs much of the sulfur from the charge, comprises about 20% by mass of iron production. It is estimated that approximately 14 million metric tons (15.5 million US tons) of blast furnace slag is produced annually in the United States.

Coal Bottom Ash/Boiler Slag

Coal bottom ash and boiler slag are coarse, granular, incombustible materials that are collected from the bottom of furnaces that burn coal. The majority of coal bottom ash and boiler slag are produced at coal-fired electric utility generation stations, although bottom ash and boiler slag are produced by industrial and institutional coal-fired boilers and also from independent coal-burning electric generation facilities. The type of bottom ash or boiler slag produced depends on the type of furnace. The American Coal Ash Association (ACAA) estimates that during 2006, the utility industry in the United States generated 1.8 million metric tons (2.0 million US tons) of boiler slag.

Coal Fly Ash

The fly ash (fly ash) produced from burning pulverized coal in a coal-fired boiler is a fine-grained, powdery particulate material that is carried off in the flue gas and usually collected by means of electrostatic precipitators, baghouses, or mechanical collection devices such as cyclones. In 2006, the American Coal Ash Association (ACAA) reported that 65.7 million metric tons (72.4 million US tons) of coal fly ash were produced.

Flue Gas Desulfurization (FGD) Scrubber Material

The burning of coal produces sulfur dioxide (SO_2) gas. The 1990 Clean Air Act and subsequent amendments

regulate SO_2 emissions from burning coal. Coal-fired power plants installed flue gas desulfurization (FGD) technology for reducing SO_2 emissions. The two most common scrubber technologies used today are wet or dry systems. Wet FGD systems introduce a spray of alkaline sorbent consisting of lime or limestone (primarily limestone) into the exhaust gas. The alkali reacts with the SO_2 gas to form calcium sulfite (CaSO_3) or calcium sulfate (CaSO_4) that is collected in a liquid slurry form. The calcium sulfite or sulfate is allowed to settle out, and the majority of the water is recycled. The settled material called FGD scrubber material or scrubber sludge is an off-white slurry with a solid content in the range of 5–10%. In 2006, FGD scrubber systems at coal-fired power plants generated approximately 27.4 million metric tons (30.2 million US tons) of material, or 24.2% of all coal combustion products.⁽⁷⁾ Due to an expected increase in scrubbing applications, the amount of FGD produced is expected to grow during the next few decades.

Foundry Sand

Foundry sand is a high-quality silica sand that is used to form molds for ferrous (iron and steel) and nonferrous (copper, aluminum, brass, etc.) metal castings. The raw sand is normally of higher quality than typical bank run or natural sands used in construction. Of the 5.5–9 million metric tons (6–10 million US tons) of spent foundry sand generated annually, less than 15% is recycled.

Kiln Dust

Kiln dusts (CKD) are fine by-products of Portland cement and lime high-temperature rotary kiln production operations that are captured in the air pollution control dust collection system (e.g., cyclones, electrostatic precipitators, and baghouses). Approximately 12.9 million metric tons (14.2 million US tons) of CKD are produced annually.

Mineral Processing Waste

Mineral processing wastes are referred to in the Resource Conservation and Recovery Act (RCRA) as wastes that are generated during the extraction and beneficiation of ores and minerals. These wastes

can be subdivided into a number of categories: waste rock, mill tailings, coal refuse, wash slimes, and spent oil shale. The mining and processing of mineral ores results in the production of large quantities of residual wastes that are, for the most part, earth- or rock-like in nature. It is estimated that the mining and processing of mineral ores generate approximately 1.6 billion metric tons (1.8 billion US tons) of mineral processing waste each year in the United States.

Municipal Solid Waste (MSW) Combustor Ash

Municipal solid waste (MSW) combustor ash is the by-product that is produced during the combustion of municipal solid waste in solid waste combustor facilities. Approximately 29.5 million metric tons (32.5 million US tons) of solid waste is combusted annually at approximately 160 municipal waste combustor plants in the United States, existing and under construction.

Nonferrous Slag

Nonferrous slags are produced during the recovery and processing of nonferrous metal from natural ores. The slags are molten by-products of high-temperature processes that are primarily used to separate the metal and nonmetal constituents contained in the bulk ore. When cooled, the molten slag converts to a rock-like or granular material. The following nonferrous slags are included in these guidelines: copper, nickel, phosphorus, lead, lead-zinc, and zinc. Approximately 3.6 million metric tons (4 million US tons) each of copper and phosphorus slag are produced each year in the United States, while the annual production of nickel, lead, and zinc slags is estimated to be in the range of 0.45–0.9 million metric tons (0.5–1.0 million US tons).

Quarry By-Products

Processing of crushed stone for use as construction aggregate consists of blasting, primary and secondary crushing, washing, screening, and stockpiling operations. Quarry by-products are produced during crushing and washing operations. There are three types of quarry by-products resulting from these

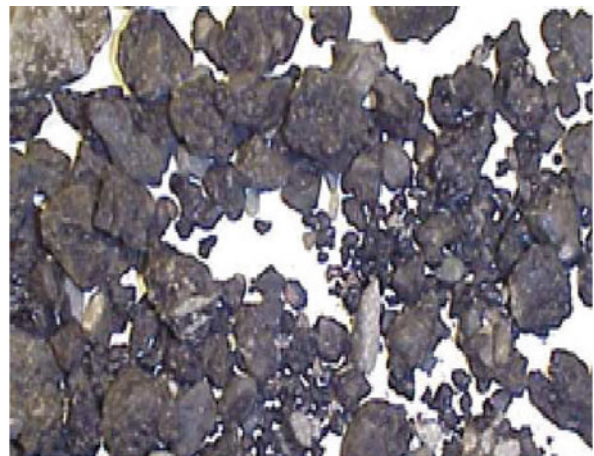
operations: screenings, pond fines, and baghouse fines. It is estimated that at least 159 million metric tons (175 million US tons) of quarry by-products are being generated each year, mostly from crushed stone production operations. As much as 3.6 billion metric tons (4 billion US tons) of quarry by-products have probably accumulated.

Reclaimed (or Recycled) Asphalt Pavement (RAP)

Reclaimed (recycled) asphalt pavement (RAP) is the term given to removed and/or reprocessed pavement materials containing asphalt and aggregates. These materials are generated when asphalt pavements are removed for reconstruction, resurfacing, or to obtain access to buried utilities. When properly crushed and screened, RAP consists of high-quality, well-graded aggregates coated by asphalt cement. It is estimated that as much as 80 million metric tons (88 million US tons) of RAP may be produced each year in the United States. A sample of RAP is shown in [Fig. 1](#).

Reclaimed Concrete Material

Over two billion metric tons (2.2 billion US tons) of aggregate are produced every year in the USA. Production of aggregate is expected to reach 2.5 billion metric tons (2.8 US tons) by 2020, which raises concerns about



Waste Materials in Construction, Utilization of. Figure 1
RAP as aggregate replacement

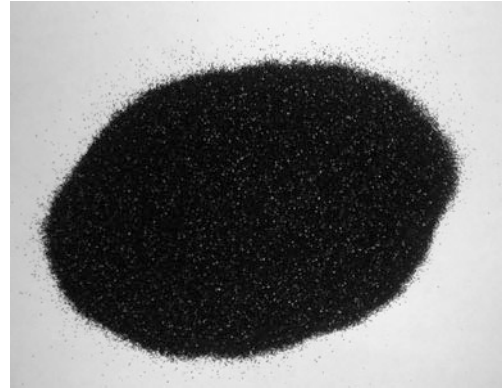
where the new aggregate will come from. One alternative to using virgin aggregate is recycled concrete aggregate. Recycled concrete aggregate (RCA) consists of high-quality, well-graded aggregates (usually mineral aggregates), bonded by a hardened cement paste. The aggregates comprise approximately 60–75% of the total volume of concrete.

Roofing Shingle Scrap

There are two types of roofing shingle scraps. They are referred to as tear-off roofing shingles, and roofing shingle tabs, also called prompt roofing shingle scrap. Tear-off roofing shingles are generated during the demolition or replacement of existing roofs. Roofing shingle tabs are generated when new asphalt shingles are trimmed during production to the required physical dimensions or from “out-of-spec” shingles. Roofing shingle tab materials that result from shingle manufacture are estimated to range between 5% and 10% of the total production. Current estimates indicate that the percentage of roofing scrap generated by activity is as follows: 87.5% is manufacturing scrap and 12.5% is tear-off roofing scrap. Approximately 10 million metric tons (11 million US tons) of asphalt roofing shingle scrap is generated each year in the United States.

Scrap Tires

Approximately 400 million tires are discarded each year by American motorists, approximately one tire for every person in the United States. Around 30 million of these tires are retreaded or reused, leaving roughly 250 million scrap tires to be managed annually. About 85% of these scrap tires are automobile tires, the remainder being truck tires. Besides the need to manage the scrap tires, it has been estimated that there may be as many as 3–4 billion tires that have accumulated over the years and are contained in numerous stockpiles. Scrap tires can be managed as a whole tire, a slit tire, a shredded or chipped tire, as ground rubber, or as a crumb rubber product. Tire rubber is also known as crumb-rubber-modifier (CRM) as it is used for modifying asphalt properties. A sample of ground tire rubber is shown in Fig. 2.



Waste Materials in Construction, Utilization of. Figure 2
Ground tire rubber

Sewage Sludge Ash

Sewage sludge ash is the by-product produced during the combustion of dewatered sewage sludge in an incinerator. Sewage sludge ash is primarily a silty material with some sand-sized particles. There are approximately 170 municipal sewage treatment plant incinerators in the United States, processing approximately 20% of the nation's sludge, and producing between 0.45 million and 0.9 million metric tons (0.5 and 1.0 million US tons) of sludge ash on an annual basis.

Steel Slag

Steel slag, a by-product of steel making, is produced during the separation of the molten steel from impurities in steel-making furnaces. The slag occurs as a molten liquid melt and is a complex solution of silicates and oxides that solidifies upon cooling.

Sulfate Waste

Fluorogypsum and phosphogypsum are sulfate-rich by-products generated during the production of hydrofluoric and phosphoric acid, respectively. Approximately 90,000 metric tons (100,000 US tons) of fluorogypsum are generated annually in the United States, mostly in Delaware, New Jersey, Louisiana, and Texas. Approximately 32 million metric tons (35 million US tons) of phosphogypsum are produced annually, mostly in central Florida, but also in Louisiana and southeastern Texas.

Applications of Waste Materials in Building Construction

Uses of waste materials is not limited to pavement applications. Huge amounts of recycled materials such as coal combustion products (CCPs), construction and demolition (C&D) materials, spent foundry sand, used tires, and slags are potentially usable in building construction applications. Reusing and recycling C&D materials onsite can reduce material hauling and disposal costs. In some cases, these materials have performance benefits over virgin materials. For example, fly ash and slag cement are used in concrete applications and can be used as partial replacements for cement, which helps in producing stronger, longer-lasting concrete. Recycled portland cement concrete has a very wide range of applications in construction. Concrete aggregate can include components from bottom ash, foundry sand, crushed concrete, and blast furnace slag. Coal bottom ash is used as bedding material for green roofs and is lighter than the alternatives. Tire shreds provide more effective drainage around building foundations in comparison with gravel and also put less pressure on the foundation walls.

Structural fills are engineered fills constructed in layers and compacted using coal fly ash, bottom ash, slag, and spent foundry. Spent foundry can be used in place of natural soil as base material for the building sites and because it does not freeze as readily as most soils, this can extend construction season. In backfills which surround the building's foundation to support it and provide drainage, waste materials can be used such as scrap tires which provide superior drainage, insulation, and wall pressure relief. Masonry blocks used in building are made from slag and aggregate, for which slag cement, fly ash, silica fume can substitute partially for cement and bottom ash; blast furnace slag and recycled concrete aggregate can be used instead of virgin aggregate.

Recycled materials are used to make sidewalks. Used tires are recycled to create rubberized sidewalks, and industrials by-products are used for concrete sidewalks. Recycled asphalt shingles and recycled asphalt pavement are used to make asphalt sidewalks. Flooring applications can also benefit from waste and recycled materials such as used tires and fly ash in carpet

backing. Salvaged lumber or recycled wood is used for wood flooring and blast furnace slag or fly ash in flooring tile. For insulation, air-cooled blast furnace slag can be used to produce mineral or rock wool insulation. There are numerous applications of waste or recycled materials in construction and new applications are emerging every day out of ongoing research on this topic.

Applicability of Current Specification to Recycled Materials

In general, current specifications on construction materials do not reflect the current level of knowledge concerning desirable properties for construction applications. It is also reasonable to state that current testing does not provide all necessary property characterization to ensure field performance. Current testing may or may not be the best testing to characterize recycled material properties. The discussions with state engineers on the applicability of current specifications on recycled materials were focused on the following points:

- Are current requirements reflecting the needs of state agencies in material quality?
- Can an agency always use current specifications to specify the best quality recycled material?
- If new specification requirements are to be introduced, would the new requirements satisfy the performance-related properties of the applications?
- The need for additional testing must be justified by relating proposed testing to performance. For example, agencies express concerns that current relations between recycled aggregate testing and pavement performance are not sufficient and further correlations are needed before making changes to current specifications.
- The ease and affordability of testing to frequently ensure consistency of aggregate quality.
- Complicated and time-consuming testing is likely to be conducted less frequently and will not be sufficient to characterize recycled materials for their variability.

Can recycled materials fully replace virgin materials with respect to current specifications? For example, is RAP aggregate?

Future Directions

There is a need to evaluate compliance and consistency of recycled material properties with respect to current specification requirements and to assess property-based requirements of recycled materials. Officials are receptive to pursuing changes in the way they specify recycled materials. It is agreeable that current specifications are still outdated for virgin materials and do not reflect the current level of knowledge concerning desirable properties for pavement applications and do not provide all necessary property characterization to ensure field performance. Current procedures may or may not be the best testing to characterize the properties of recycled materials. There have been discussions regarding added testing procedures (to measure additional properties) in the current specifications to ensure the applicability of specifications on recycled materials. An example is asphalt-CRM binder stability. The concern regarding the relation between the suggested testing and performance has two aspects: If those relations are valid, then there is a need for a change in current specifications. If those relations do not exist, then marginal materials, including local sources that are poor in those properties, should be used successfully in most civil/construction applications.

Specific significant testing procedures that are well correlated to performance are difficult to implement on recycled materials. An example is aggregate compressive strength, ASTM C-170 Standard Test Method for Compressive Strength of Dimension Stone. The test cannot be considered for recycled aggregate sources because it requires a cubic-cut sample of the aggregate source of more than 5.0 cm each dimension. However, an alternative test, the point load test (ASTM D5731), uses much smaller sample and can provide reasonable accuracy for measuring the same property.

The comparison between the qualities of local materials vs. that of recycled materials is a sustainability issue. Aggregate is an example. There is a potential that local aggregate sources that are rejected for high-class pavement mixes can be used for local and low volume road applications. Aggregate quality is tied to the design process. The use of recycled materials, recycled asphalt pavement (RAP) for example, has been successful in same applications. Cost is a major factor in selecting aggregate

sources. Life-Cycle-Cost analysis is a useful tool in determining the use of local aggregate vs. recycled materials applications. Having better quality aggregate would allow designing pavement for longer life, an aspect of sustainability.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 0846861. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Bibliography

Primary Literature

1. Recycling Centers website. <http://www.recyclingcenters.org>
2. <http://www.dosomething.org/whatsyourthing/Environment>
3. Environmental Protection Agency website. <http://www.epa.gov>
4. The Construction Industry Compliance Assistance Center. <http://www.cicacenter.org/>
5. Recycling, Wikipedia. <http://en.wikipedia.org/wiki/Recycling>
6. Kandhal PS, Mallick RB (1997) Pavement recycling guidelines for state and local governments participant's reference book. Publication No. FHWA-SA-98-042, Federal Highway Administration, Washington, DC
7. Wright FG Jr (2002) Formal policy on the use of recycled materials. Federal Highway Association, Washington, DC
8. Recycled Materials Resource Center, User guidelines for byproducts and secondary use materials in pavement construction. <http://www.rmrc.unh.edu/>
9. Environmental Protection Agency, Using recycled industrial materials in building
10. Eighmy TT, Chesner WH (2000) Framework for evaluating use of recycled materials in the highway environment. Publication No. FHWA-RD-00-140, Federal Highway Administration, Washington, DC

Books and Reviews

- Ahmed I (1993) Use of waste materials in highway construction. Noyes Data Corporation, Park Ridge
- Boden F, Azaroual M, Piantone P (2000) Forecasting the behaviour of long-term municipal solid waste incineration bottom ash: rapid tests combined. In: Woolley GR, Goumans JJM, Wainwright PJ (eds) Waste materials in construction. Proceedings of the international conference on the science and engineering of recycling for environmental protection, 1–2 June 2000. Elsevier, Amsterdam

- Bossinik BAG, Brouwers HJH (1996) Construction waste: quantification and source evaluation. *J Constr Eng Manag ASCE* 122(1):55–60
- Calkins M (2009) *Materials for sustainable sites: a complete guide to the evaluation, selection and use of sustainable construction materials*. Wiley, Hoboken
- Carpenter AC, Gardner KH (2007) Recycled material highway construction environmental assessment: life cycle based risk assessment of recycled materials in roadway construction. RMRC Project No. 43 final report, Recycled Materials Resource Center
- Carter A, Gardiner MS (2007) Indirect tension relaxation test to evaluate the effect of the addition of RAP to HMA mixes. *J Mater Civ Eng ASCE* 19(3):219–226
- Chandra S (1997) *Waste materials used in concrete manufacturing*. Noyes Data Corporation, Westwood
- Chen D-H, Hugo F (2001) Comparison of two pavement rehabilitation strategies. *J Transp Eng ASCE* 127(1):47–58
- Cross SA (2002) Determination of N_{Design} for CIR mixtures using the superpave gyratory compactor. RMRC Research Project No. 15 final report, Recycled Material Resource Center
- Federal Highway Administration (1993) A study of the use of recycled paving material – report to congress. Publication No. FHWA-RD-93-147, Federal Highway Administration, Washington, DC
- Federal Highway Administration (2004) *Transportation applications of recycled concrete aggregate*. FHWA State of the Practice National Review. Federal Highway Administration, Washington, DC
- Gidley JS, Sack WA (1984) Environmental aspects of waste utilization in construction. *J Environ Eng ASCE* 110(6): 1117–1133
- Goumans JJM, van der Sloot HA, Aalbers ThG (eds) (1994) *Environmental aspects of construction with waste materials*. Proceeding of the international conference on environmental implications of construction materials and technology developments, Maastricht, 1–3 June 1994. Elsevier, Amsterdam
- Haselbach L (2008) *The engineering guide to LEED—new construction: sustainable construction for engineers*. McGraw-Hill, New York
- Horvath A (2004) A life-cycle model and decision-support tool for selecting recycled versus virgin material for highway applications. RMRC Project No. 23 final report, Recycled Material Resource Center
- Inyang HI (2003) Framework for recycling of wastes in construction. *J Environ Eng ASCE* 129(10):887–898
- Kelly T (1998) Crushed cement concrete substitution for construction aggregates – a material flow analysis, U.S. Geological Survey Circular 1177. <http://greenwood.cr.usgs.gov/pub/circulars/c1177/index.html>
- Kim JJ, Lee HD, Jahren CT, Heitzman M (2010) Long-term field performance of cold-in-place recycled road s in Iowa. *J Perform Constr Facil ASCE* 24(3):265–274
- Kim Y, Lee HD, Heitzman M (2007) Validation of new mix design procedure for cold-in-place recycling with foamed asphat. *J Mater Civ Eng ASCE* 19(11):1000–1010
- Lee KW, Brayton TE, Huston M (2002) Development of performance based mix design for cold-in-place recycling (CIR) of bituminous pavements based in fundamental properties. Publication No. FHWA-CIR-02-01, Federal Highway Administration, Washington, DC
- Mallick RB, Mathisen PP, FitzPatrick MS (2002) Opening the window of sustainable development to future civil engineers. *J Prof Issue Eng Educ Pract ASCE* 128(4): 212–216
- McDaniel RS, Soleymani H, Shah A (2002) Use of reclaimed asphalt (RAP) under superpave specifications. Publication No. FHWA / IN/JRTP-2002/6
- Melton JS, Morgan T, Evaluation of tests for recycled material aggregates for use in unbound applications. RMRC Project 6 final report, Recycled Materials Resource Center
- Naik TR (2008) Sustainability of concrete construction. Practice periodical on structural design and construction. *Am Soc Civ Eng* 13(2):98–103
- Nassar K, Nassar W (2006) Reclaimed asphalt pavement detection and quality determination practice periodical on structural design and construction. *Am Soc Civ Eng* 11 (3):171–176
- Polley C, Cramer SM, de la Cruz RV (1998) Potential for using waste glass in portland cement concrete. *J Mater Civ Eng ASCE* 10(4):210–219
- Pratt E (1993) Current uses and evaluation of recycled materials in highway construction: overview of northeastern states. Northeastern Recycling Council
- Rajendran S, Gambatese JA (2007) Solid waste generation in asphalt and reinforced concrete roadway life cycles. *J Infrastruct Syst ASCE* 13(2):88–96
- Rubber Manufacturers Association (2009) *Scrap tire markets in the United States*, 9th biennial report. Rubber Manufacturers Association, Washington, DC
- Saeed A (2008) Performance-related tests of recycled aggregates for use in unbound pavement layers. NCHRP Report 598, National Cooperative of Highway Research Program
- Schimmoller VE et al (2000) Recycled materials in European highway environments: uses, technologies and policies. Publication No. FHWA-PL-00-025, Federal Highway Administration, Washington, DC
- Shen D-H, Du J-C (2005) Application of gray relational analysis to evaluate HMA with reclaimed building materials. *J Mater Civ Eng ASCE* 17(4):400–406
- Shen J, Amirghanian S, Miller JA (2007) Effects of rejuvenating agents on superpave mixtures containing reclaimed asphalt pavements. *J Mater Civ Eng ASCE* 19(5):376–384
- Sobhan K, Mashnad M (2000) Fatigue durability of stabilized recycled aggregate base course containing fly ash and waste-plastic strip reinforcement. Recycled Material Resource Center
- Sorvari J (2003) By-products in earth construction: environmental assessments. *J Environ Eng ASCE* 129(10):899–899

- Taha B, Nouno G (2009) Utilizing waste recycled glass as sand/cement replacement in concrete. *J Mater Civ Eng ASCE* 21(12):709–721
- Tao M, Zhang Z, Wu Z (2008) Simple procedure to assess performance and cost benefits of using recycled materials in pavement construction. *J Mater Civ Eng ASCE* 20(11):718–725
- Tay J, Hong S, Show K (2000) Reuse of industrial sludge as pelletized aggregate for concrete. *J Environ Eng ASCE* 126(3):279–287
- Venner M (2004) Environmental stewardship practices, procedures, and policies for highway construction and maintenance. NCHRP Project 25-25(4), National Cooperative Highway Research Program
- Wang Y (2009) Sustainability in education. *J Prof Issues Eng Pract ASCE* 135(1):21–30
- Willburn DR, Goonan TG (1998) Aggregates from natural and recycled sources, economic assessment for construction applications – a material flow analysis, U.S. Geological Survey Circular 1176. <http://greenwood.cr.usgs.gov/pub/circulars/c1176/c1176.html>
- Young HM, Sellasie K, Zeroka D, Sabnis G (2003) Physical and chemical properties of recycled tire shreds for use in construction. *J Environ Eng ASCE* 129(10):921–929
- Zapata P, Gambatese JA (2005) Energy consumption of asphalt reinforced concrete pavement materials and construction. *J Infrastruct Syst ASCE* 11(1):9–20
- Zhang W, Ingham JM (2010) Using recycled concrete aggregates in New Zealand ready-mix concrete production. *J Mater Civ Eng ASCE* 22(5):443–450

Related Websites

- Aggregate Research Institute. <http://www.aggregateresearch.com/>
- Asphalt Recycling and Reclaiming Association. <http://www.arra.org/>
- Benefits of Recycling. <http://www.benefits-of-recycling.com/recyclingstatistics.html>
- Center for Environmental Excellence by AASHTO website. <http://environment.transportation.org/>
- Environmental Protection Agency, Waste-resource conservation. <http://www.epa.gov/epawaste/conserve/index.htm>
- Federal Highway Administration recycling webpage. <http://www.fhwa.dot.gov/pavement/recycling/index.cfm>
- National Demolition Association. <http://www.demolitionassociation.com/>
- Pavement Recycling Systems. <http://www.pavementrecycling.com/>
- Recycled Materials Resource Center. <http://www.rmrc.unh.edu/>
- Recycling Centers. <http://www.recyclingcenters.org>
- Recycling Today, The Industry's News and Resource Center. <http://www.recyclingtoday.com/>
- Sustainable Sources, Construction waste recycling. <http://constructionwaste.sustainable-sources.com/>
- The Construction Industry Compliance Center. <http://www.cicacenter.org/cdrecycling.html>

Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery

HELMUT RECHBERGER

Institute of Water Quality, Resource and Waste Management, Vienna University of Technology, Vienna, Austria

Article Outline

Glossary
 Definition of the Subject and Its Importance
 Introduction: Material Balances and Entropy Evaluation
 Description of Waste-to-Energy (WTE)
 Statistical Entropy Analysis (SEA)
 Application of SEA to Waste-to-Energy
 Relationship Between SEA and Environmental Impact
 Future Directions
 Bibliography

Glossary

- Statistical entropy** Developed by C.E. Shannon as part of his information theory and used as a measure for distributions in statistics, formally similar to the thermodynamic entropy of mixing.
- Waste-to-energy (WTE)** Combustion of municipal solid waste (MSW) with energy recovery.
- Material flow analysis (MFA)** Systematic assessment of the flows and stocks of materials within a system, defined in space and time.
- Statistical entropy analysis (SEA)** Entropy-based method, based on material flow analysis, for quantifying the concentrating effect of a process or system.

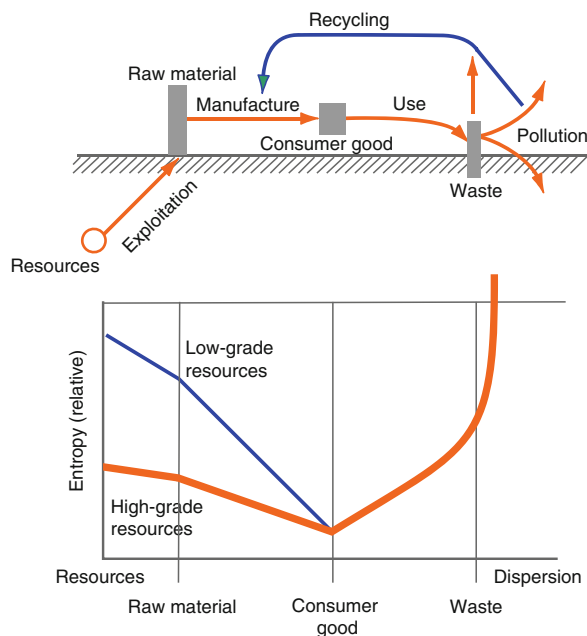
Definition of the Subject and Its Importance

The turnover of materials used in a national economy has been described as the consumption of resources (low-entropy materials) that are used and transformed into wastes (high entropy materials); therefore, the economy is viewed as an entropy-producing process [1]. In a recycling economy, entropy generation must

be kept low and waste management should transform high-entropy wastes into low-entropy recycled products that can reduce the use of primary resources. Entropy reduction is achieved by the concentrating of materials. Evaluation of the materials balances, by means of statistical entropy analysis, within state-of-the-art incinerators shows that waste-to-energy is a major contributor to a sustainable materials management. It is shown that waste incineration results in significant concentrating of several substances.

Introduction: Material Balances and Entropy Evaluation

Stumm and Davis [2] described the life cycle of copper qualitatively by using the metric of entropy. They showed that entropy is reduced when consumer products are made from raw materials and entropy increases after the products are used and become waste (Fig. 1). Later, Rechberger and Graedel [3] applied the concept of statistical entropy to the global copper cycle and quantified the entropy trend (Fig. 2).



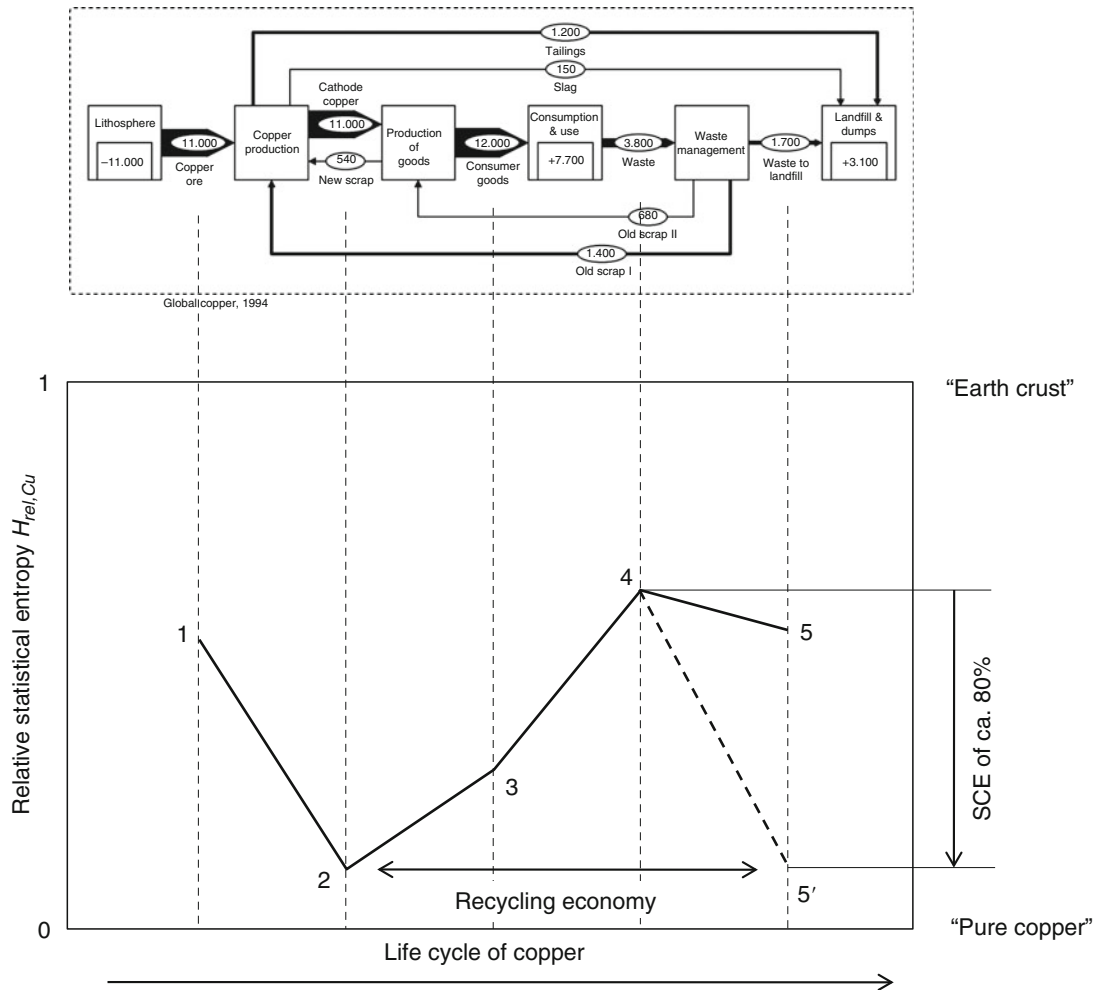
Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Figure 1
The life cycle of copper and the change in its entropy [2]

The shape of the Relative Statistical Entropy (RSE) curve of Fig. 2 can be explained as follows: One tracks the life cycle of a batch of copper (e.g., 1 kg) and quantifies the state of this amount of copper by applying statistical entropy. There are two extreme states for the copper batch: either it appears as a 100% pure item, e.g., as a copper bar of 1 kg, which translates into a RSE of zero, or it is completely dissipated and contributes to the mean concentration in the earth's crust (ca. 70 mg/kg), which is defined as the maximum entropy state (RSE of 1). Any other state has to lie between these two extremes. The life cycle starts with copper ore (Fig. 2, point 1), which has a mean concentration of 0.5%. The "copper production" process comprises several concentrating and beneficiation steps resulting in high-grade cathode copper having a purity of 99.9% or higher (step 1 → 2 in Fig. 2). However, point 2 lies above the x-axis (RSE = 0) because extracting copper from its ore is not 100% efficient; there are some waste by-products (tailings, slag) and these account for the distance between point 2 and the zero-line in Fig. 2.

The process "production of goods" requires that copper metal be mixed or combined with other materials to form consumer products having copper concentrations between 0.01% and 100%; this results in an entropy increase (step 2 → 3 in Fig. 2). These products are then used and eventually become waste (step 3 → 4) with a typical copper concentration of <1% (MSW, waste electrical and electronic equipment, end-of-life vehicles, construction and demolition waste, etc.).

The process of "waste management" separates copper scrap and this results in an entropy decrease (step 4 → 5 in Fig. 2). The decrease is rather moderate because much of the copper becomes embedded in the stock of the process "consumption and use" that represents a future potential for recycling (and entropy decrease); because of this, the global recycling rate is rather low (about 50%).

The global total entropy trend for copper is rather balanced since point 1 (ore) and point 5 (wastes to landfill) have similar values. However, in a fully developed recycling economy where waste becomes a high-grade resource again, the life cycle of copper would end at point 5' in Fig. 2. This means that waste management needs to realize a substantial reduction in entropy, posing a similar challenge as the



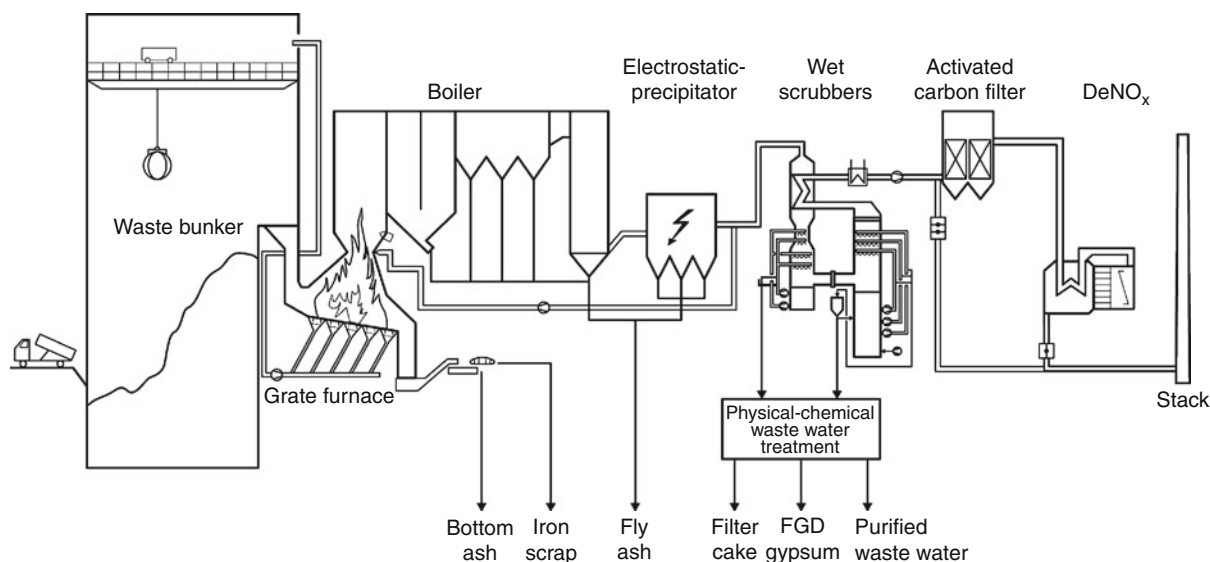
Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Figure 2
Global production and use of copper (upper diagram) and corresponding change in its relative statistical entropy [3]

process from step 1 → 2 (Fig. 2), i.e., the production of pure cathode copper from copper ore. The latter is achieved by a series of processes, such as grinding, flotation, roasting, smelting, extraction, and refining. Therefore, it is straight forward to postulate that the path from 4 → 5' (Fig. 2) will require a similar technological effort as the production path from 1 → 2. In other words, the realization of a substantial recycling economy will require high-tech waste management. At present, these high-tech technologies have not yet been introduced to waste management. Notable exceptions are state-of-the-art incinerators and a

trend to sophisticated recycling technologies such as near-infrared spectroscopy detection and sorting equipment.

Description of Waste-to-Energy (WTE)

A waste-to-energy facility can be divided into the actual combustion process, the boiler and power generation turbine, and the air and water pollution control system. The choice for the combustion technology is mainly determined by the physical (particle size distribution, density, aggregate state, water content, etc.) and



Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Figure 3
Modern WTE facility with state-of-the-art air pollution control (APC) system

chemical (elemental composition, calorific value, ash content, etc.) characteristics of the waste. A widely used method for the incineration of municipal solid waste (MSW) is the so-called mass-burn grate combustion (Fig. 3).

In a typical mass-burn facility [4], waste is stored in a fully contained waste bunker for periods up to 1 week. Combustion air is taken from the bunker thus preventing odor outside the facility. An operator-controlled or automated crane mixes incoming wastes to provide some homogeneity with regard to calorific value and loads the waste into a hopper feeding a water-cooled gravity chute that serves as an air seal between bunker and furnace. A horizontal ram feeder at the bottom of the chute pushes the waste onto the furnace grate where it is combusted. The grate consists of moving and fixed elements (rows) that guarantee controlled transportation, mixing of the waste, and equal distribution across the grate. At the end of the grate, the bottom ash is discharged into a water basin, where it is cooled and then conveyed to the bottom ash bunker.

The flue gas leaving the combustion chamber contains high amounts of particulates ($3\text{--}5\text{ g/m}^3$), acids such as SO_2 (sulfur dioxide), HCl (hydrochloric acid), HF (hydrofluoric acid), and NO_x (nitrogen oxides) that

require extensive flue gas treatment. Particulates are removed in electrostatic precipitators (ESP) or more commonly in bag house filters (also referred to as fabric filters). ESPs use electric forces to charge and move particulate matter in a flowing gas stream to a collecting surface. They are more robust toward temperature changes, have lower pressure drops, and are easy to maintain. Bag house filters (particulate removal by filtration on fiber surfaces) achieve high removal efficiencies for fine particles ($<5\text{ }\mu\text{m}$, $>99.9\%$).

Various reagents are used in the APC system of WTEs, such as lime (CaO) or lime hydrate (Ca(OH)_2), to remove sulfur dioxide and hydrogen chloride, and activated carbon, to remove mercury and dioxins. Acids can be very efficiently and separately removed by a two-stage wet scrubber system. HCl and HF are physically absorbed in the first acid scrubber stage at pH1. The low pH prevents absorption of SO_2 . The scrubber liquid is water, and no neutralization agent has to be used. Also mercury chlorides are removed from the flue gas at this stage. During the second stage SO_2 is chemically absorbed using either CaCO_3 , Ca(OH)_2 , or NaOH (sodium hydroxide) for neutralization at pH7. The final product of the sulfur removal is gypsum, which is dewatered and either

recycled in the gypsum industry or landfilled. Part of the scrubber water is continuously purified by an on-site physical-chemical wastewater treatment process. Neutralization sludge is precipitated and dewatered in a filter press. Chloride is discharged with the treated wastewater.

Hydrogen chloride (HCl) can be purified and concentrated to over 30% for recycling. Alternatively, the acid can be neutralized with NaOH, and sodium chloride brine can be recycled. NO_x is reduced to N₂ and H₂O by injecting NH₃ (ammonia). The reaction requires high temperatures (900–1,000°C). A ceramic catalyst (based on TiO₂ and V₂O₅) reduces the temperature need to 200–300°C and enhances the efficiency of the reaction from 70% to 80% up to 90%. With proper design, the catalyst can also be utilized to oxidize PCDD/F. Sometimes activated carbon filters are employed as a final stage to adsorb organic compounds (e.g., PCDD/F), metallic mercury (Hg⁰), and other trace pollutants passing the previous air pollution control (APC) stages. State-of-the-art combustion and APC systems destroy organic and remove inorganic pollutants with high efficiency. Various full-scale plants demonstrate the reliability of these technologies; they operate at emissions way below the stringent emission limits enacted, e.g., by the European Union [5].

An ongoing field of research and development with regard to WTE is how to produce bottom ash and filter residues that have improved landfill properties with regard to composition and emissions; and residues that can be utilized as secondary resources. This can be achieved either by after-treatment of bottom and/or fly ashes in thermal and/or chemical processes or by designing new technologies that are not necessarily based on the grate furnace technology.

Bottom ashes can also be further treated by mechanical processes (sieving, screening) and magnetic separation of iron scrap. Adequate thermal after-treatment of bottom ash results in three products: (1) a silicate product that can be utilized for construction purposes; (2) metals consisting mainly iron, copper, and other lithophilic metals (metals of low vapor pressure); and (3) a concentrate of atmophilic metals (metals of high vapor pressure). The last two fractions can be recycled in the metal industry.

Immobilization of leachable substances can also be achieved by solidification/stabilization using additives such as cement that physically and/or chemically immobilize hazardous substances in ashes. Some new incineration concepts combine pyrolysis and high-temperature processes with conventional APC as described in other entries of this encyclopedia. The goals of these technologies are to produce residues with better qualities with respect to disposal and recycling than mass burn systems. None of these new technologies have experienced a breakthrough. Up to now, they could not prove yet that they are able to reach the same or higher goals than traditional MSW incineration at lower costs and at the same reliability. Today, WTE combined with advanced APC represents a reliable, robust, and compared to other available options environmentally sound technology to dispose of combustible and hazardous wastes. Further development should be focused on producing more residues that can be recovered in an environmentally sound manner, and to use thermal processes to turn waste management into an integrated part of a sustainable materials management.

Table 1 provides the mass balance of a WTE facility consisting of mass-burn combustion, ESP, two-stage wet scrubber system, de-NO_x reactor, activated carbon filter, and chemical-physical wastewater treatment. Table 2 shows the respective transfer coefficients for selected metals. Both were determined by an intensive measurement campaign at a full-scale facility over 1 week [6].

Statistical Entropy Analysis (SEA)

The methodological framework of SEA is described in [7] and [8]. SEA employs statistical entropy to quantify the distribution of substances or elements in a process. In statistics, the entropy is used as a metric to describe distributions of any kind. Rechberger showed that the input as well as the output of a process (e.g., a WTE facility) can be described by distributions of substances as a result of a material flow analysis (MFA). Figure 4 and Table 3 show an example, the cadmium balance of a WTE facility.

The distribution of the input is rather simple since over 99% of the Cd enters the process via the material flow “MSW” that contains a concentration of

Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Table 1 Normalized mass balance of a state-of-the-art WTE facility [6]

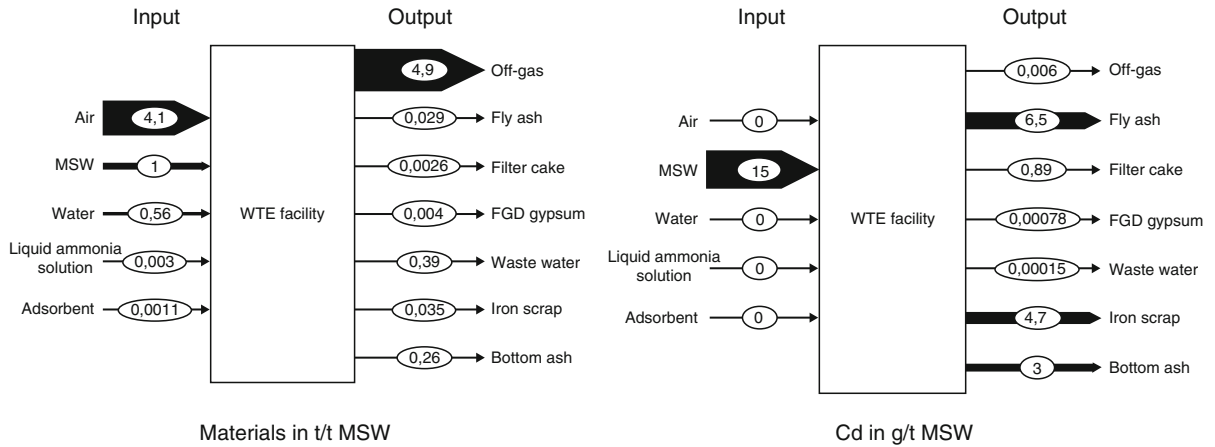
Input	t/t MSW	Output	t/t MSW
MSW	1.0	Bottom ash	0.26
Adsorbent	0.0011	FGD-gypsum	0.0040
Liquid ammonia solution	0.0030	Filter cake	0.0026
Air	4.1	Fly ash	0.029
Water	0.56	Iron-scrap	0.035
		Off-gas	4.9
		Purified waste water	0.39
Total	5.7		5.7

Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Table 2 Transfer coefficients (%) of a WTE plant [6]

	Bottom ash %	Fly ash %	Iron-scrap %	FGD-gypsum %	Filter cake %	Pur. WW %	Off-gas %
Ag	83	16	0.001	0.0062	0.48	0.06	0.001
As	70	28	0.17	0.23	0.68	0.09	0.6
Cd	20	43	31	0.0052	5.9	0.01	0.04
Co	70	9	20	0.054	0.09	0.02	0.005
Cr	83	4	13	0.0006	0.032	0.0007	0.0006
Cu	96	0.9	3	0.0005	0.03	0.00005	0.00002
Fe	40	1.2	58	0.00026	0.08	0.00031	0.001
Hg	5	4.6	0.001	3.27	83	0.3	3.9
Mn	84	12	3.8	0.00032	0.16	0.034	0.0002
Ni	83	2	15	0.0022	0.027	0.001	0.0003
Pb	74	25	0.0049	0.0078	0.8	0.0006	0.0007
Sb	41	55	2.3	0.01	1.6	0.004	0.003
Sn	76	23	0.001	0.014	0.57	0.009	0.01
V	81	11	7.6	0.0026	0.39	0.02	0.005
Zn	72	26	0.6	0.011	1.27	0.006	0.0006

15 mg/kg MSW. The output distribution is more complex. Cadmium is partitioned among the outputs as follows: Bottom ash, 20% (11 mg/kg); iron scrap, 31% (130 mg/kg); fly ash, 43% (220 mg/kg); filter cake, 5.9% (340 mg/kg); FGD gypsum, 0.0052%

(0.2 mg/kg); purified waste water, 0.01% (0.0038 mg/kg); and off-gas, 0.04% (0.0012 mg/kg). In other words, cadmium is transferred mostly to some output flows (fly ash, iron scrap) and at different concentrations in the various outputs. By applying the



Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Figure 4
Materials and Cd balance for a WTE facility. The WTE facility can be conceived as a transformer of distributions (input → output)

Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Table 3 Result of an MFA for the flow of cadmium (Cd) through a WTE facility (material flows (M_i), concentrations (c_i), and resulting substance flows (X_{ij}))

Input	M_i	c_i	X_i		Output	M_i	c_i	X_i	
	t material/ t MSW	kg Cd/kg material	kg Cd/t MSW			t material/ t MSW	kg Cd/kg material	kg Cd/t MSW	
MSW	1.0	0.000015	0.015	99.9	Bottom ash	0.26	0.000011	0.003	20
Adsorbent	0.0011	0	0	0	FGD-gypsum	0.004	0.0000002	0.00000078	0.0052
Liquid ammonia solution	0.003	0	0	0	Filter cake	0.0026	0.00034	0.00089	5.9
Air	4.1	0	0	0	Fly ash	0.029	0.00022	0.0065	43
Water	0.56	0	0	0	Iron-scrap	0.035	0.00013	0.0047	31
					Off-gas	4.9	0.000000012	0.000006	0.04
					Purified waste water	0.39	0.0000000038	0.0000015	0.01
Total	5.7		0.015	100		5.7		0.015	100

following equations, it is possible to quantify whether Cd is rather concentrated or diluted during the WTE operation.

X_{ij} ... flow of substance j in material i

M_i ... flow of material i

c_{ij} ... concentration of substance j in material i

$$X_{ij} = M_i \cdot c_{ij} \quad (1)$$

with $i = 1 \dots n_I$ for input and $i = 1 \dots n_O$ for output
 m_i ... normalized mass flows

$$m_i = \frac{M_i}{\sum_{i=1}^{n_I} X_{ij}} = \frac{M_i}{\sum_{i=1}^{n_O} X_{ij}} \quad (2)$$

The statistical entropy function for the input and output distribution is expressed as follows:

$$H_j(c_{ij}, m_i) = - \sum_{i=1}^k \bar{m}_i \cdot \bar{c}_{ij} \cdot \text{ld}(\bar{c}_{ij}) \quad (3)$$

where ld is the logarithm to the base 2 and $\text{ld}(0) = 0$.

The dilution of gaseous (g) and aqueous (a) emissions in the environment (only for the output) is considered by introducing the geogenic background concentration ($c_{j,\text{geog.}}$) of a substance in air and water. The assumption is that the emission via a stack or wastewater pipe into a river is diluted to 1% (therefore the factor of 100 in Eqs. 4 and 5) above the respective geogenic background concentration. Equations 4 and 5 are approximations and only valid if $c_{ij} \gg c_{j,\text{geog.}}$. If this is not the case, a more complex formula has to be applied; for details see [7].

$$\bar{m}_i = \begin{cases} \frac{c_{ij} \cdot m_i}{c_{j,\text{geog.},g}} \cdot 100 \\ \frac{c_{ij} \cdot m_i}{c_{j,\text{geog.},a}} \cdot 100 \\ m_i \end{cases} \quad (4)$$

$$\bar{c}_{ij} = \begin{cases} c_{j,\text{geog.},g}/100 \\ c_{j,\text{geog.},a}/100 \\ c_{ij} \end{cases} \quad (5)$$

The maximum statistical entropy is expressed by

$$H_{\max,j} = \text{ld}\left(\frac{1}{c_{j,\text{geog.}}} \cdot 100\right) \quad (6)$$

$$\begin{aligned} \text{with } c_{j,\text{geog.}} &= c_{j,\text{geog.},g} & \text{when } c_{j,\text{geog.},g} < c_{j,\text{geog.},a} \\ \text{and } c_{j,\text{geog.}} &= c_{j,\text{geog.},a} & \text{when } c_{j,\text{geog.},g} > c_{j,\text{geog.},a} \end{aligned}$$

$H_{\max}$ is determined by assuming that the total substance flow ΣX_{ij} is diluted into the environmental compartment with the lowest background concentration (in the case of heavy metals, this is usually the atmosphere).

The relative statistical entropy is a normalized metric in the range [0, 1].

$$H_{\text{rel},j} = H_j/H_{\max,j} \quad (7)$$

The power of a process (e.g., WTE facility) to concentrate a substance j is then expressed by

$$\Delta H_{\text{rel},j} = H_{\text{rel},j,\text{Output}} - H_{\text{rel},j,\text{Input}} \quad (8)$$

Alternatively, the “substance concentrating efficiency” (SCE) can be defined as a specific metric, in percentage:

$$\text{SCE}_j = \frac{-\Delta H_{\text{rel},j}}{H_{\text{rel},j,\text{Input}}} \cdot 100 \quad (9)$$

The SCE is positive for concentrating processes with a maximum of 100% (the “pure” substance) and negative for diluting processes (e.g., for Hg when refuse derived fuel is used for co-combustion in a cement kiln). As an example, Table 4 provides the calculation of ΔH_{rel} for cadmium in a WTE plant.

Application of SEA to Waste-to-Energy

By applying Eqs. 1–9 to the metals shown in Table 2, the resulting SCE-values for a specific WTE facility are shown in Fig. 5.

Generally, it can be said that the SCE of state-of-the-art WTE for most metals is in the range between 10% and 20%. For volatile (atmophilic) elements, it tends to be higher since these elements are rather transferred to and concentrated in the APC residues. The highest SCE is achieved for iron. Ferrous metals are easily separated and collected in a WTE facility, by means of magnetic separation.

Following the above discussion, it is clear that SCE-values of 10–20% are not sufficient to realize a noteworthy recycling economy (see Fig. 2). In the case of global copper management SEA shows that waste management should achieve a SCE of at least 80%.

On the other hand, it should be noted that a selective separation step does not have to be restricted to iron. Intensive characterization of bottom ash [6, 9] has shown that some metals appear to a large extent in metallic form in the ash and not as oxides or other compounds. For example, Fig. 6 shows that about 50% of the copper contained in the bottom ash would be recoverable by a mechanical process; particles larger than 4 mm are considered to be detectable and removable by a combination of near-infrared spectroscopy detection and pneumatic separators. This measure would increase the SCE for Cu from 21% to 51%. A melting process, described by Zeltner and Lichtensteiger [10] and shown in Fig. 7, would increase the SCE up to 80%. This last example shows that high recovery rates, i.e., high SCE values, can be achieved. However, they require the

Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Table 4 Example: Calculation of ΔH_{rel} and the SCE of Cd in a WTE facility

n_i	$j = \text{Cd}$	M_i	c_i	X_i	m_i	$c_i \times m_i$	$ld(c_i)$	$c_i \times m_i \times ld(c_i)$
	Input	t material/ t MSW	kg Cd/kg of material	kg Cd/t MSW	kg of material/ kg Cd	–	–	–
1	MSW	1.0	0.000015	0.015	66,664	1.0	–16.0	–16.0
2	Adsorbent	0.0011	0	0	74	0.0	0.0	0.0
3	Liquid ammonia solution	0.003	0	0	200	0.0	0.0	0.0
4	Air	4.1	0	0	273,322	0.0	0.0	0.0
5	Water	0.56	0	0	37,332	0.0	0.0	0.0
	Total	5.7		0.015	377,593	1.0		–16.0
n_o		M_i	c_i	X_i	$\bar{m}_i$	$c_i \times \bar{m}_i$	$ld(c_i)$	$c_i \times \bar{m}_i \times ld(c_i)$
	Output	t material/ t MSW	kg Cd/kg material	kg Cd/t MSW	kg material/kg Cd	–	–	–
1	Bottom ash	0.26	0.000011	0.003	17,400	0.2	–16.41	–3.3
2	FGD-gypsum	0.004	0.0000002	0.00000079	265	0.000052	–22.28	–0.00116
3	Filter cake	0.0026	0.00034	0.00089	173	0.59	–11.52	–0.68
4	Fly ash	0.029	0.00022	0.0065	1,934	0.43	–12.13	–5.22
5	Iron-scrap	0.035	0.00013	0.0047	2,347	0.31	–12.89	–4
6	Off-gas (g)	4.9	1.0E-14	0.000006	4E + 10	0.0004	–46.51	–0.0186
7	Purified waste water (a)	0.39	1.0E-12	0.00000015	1E + 08	0.0001	–39.86	–0.004
	Total	5.7		0.015	4.0E + 10	1.0		–13.2
	Geogenic concentration of Cd in				H_{max}	47		
	Air ($c_{geog,g}$)	1.0E-12	kg Cd/kg air		$H_{rel,Input}$	0.34		
	Water ($c_{geog,a}$)	1.0E-10	kg Cd/kg water		$H_{rel,Output}$	0.28	ΔH_{rel}	–0.06
					SCE (%)	18		

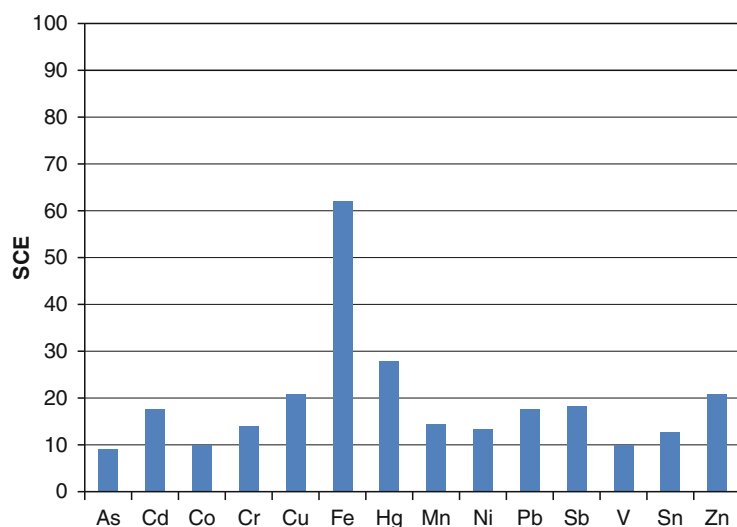
implementation of sophisticated and costly technology. Although it has not been established quantitatively as yet, it is safe to assume that SCE rates in the range of 70–90% are achievable for certain substances.

Relationship Between SEA and Environmental Impact

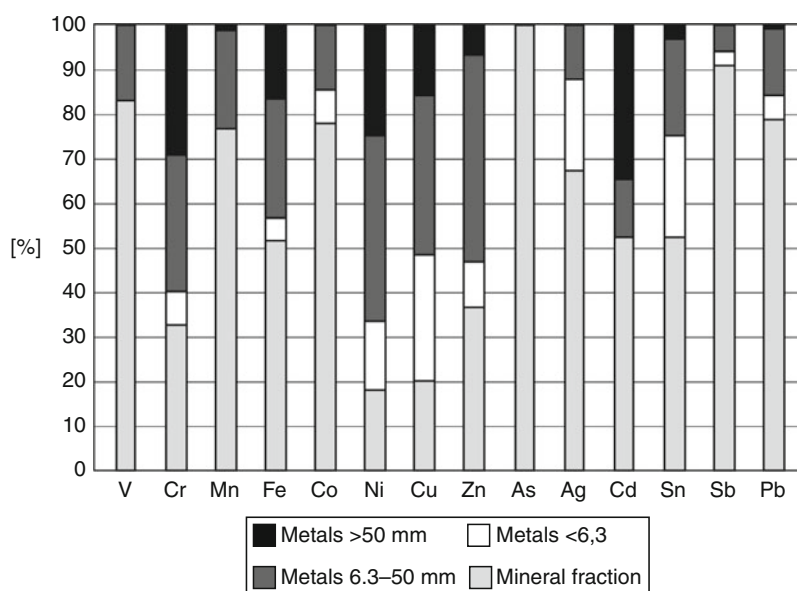
There is no scientific evidence that entropy is a direct measure of environmental impact. However, considering

the stack emissions of a WTE facility, Fig. 8 shows that the actual emissions of Cd (0.0012 mg/m^3) are in a range where they do not contribute to $H_{rel,Output}$. A tenfold increase to 0.01 mg/m^3 would result in an entropy increase by 1%. The emission standard for Cd (0.025 mg/m^3) is in the range of 1–5% which is the area the cadmium emission in stack gas would become entropy relevant.

Therefore, from the point of view of statistical entropy analysis, the emission limit of Cd has been



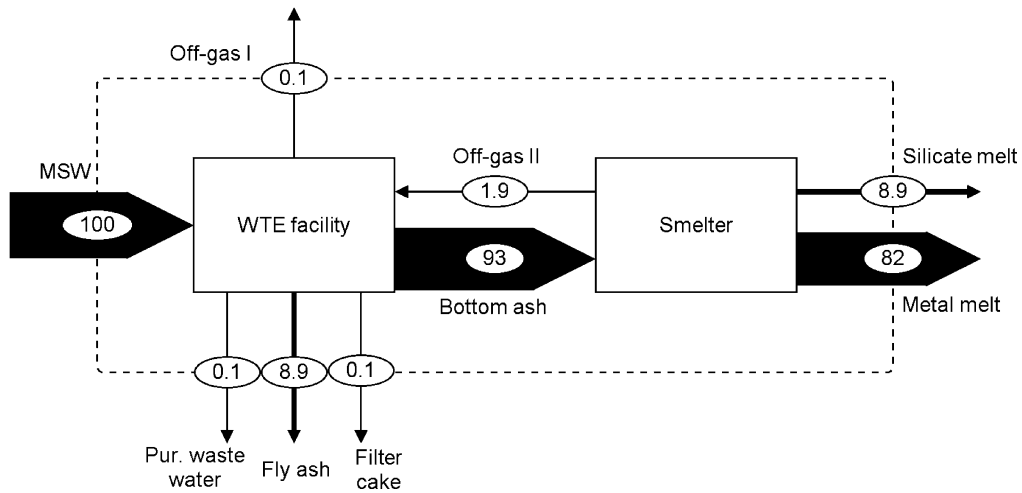
Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Figure 5
Substance concentrating efficiency (SCE in %) of a WTE facility for selected metals



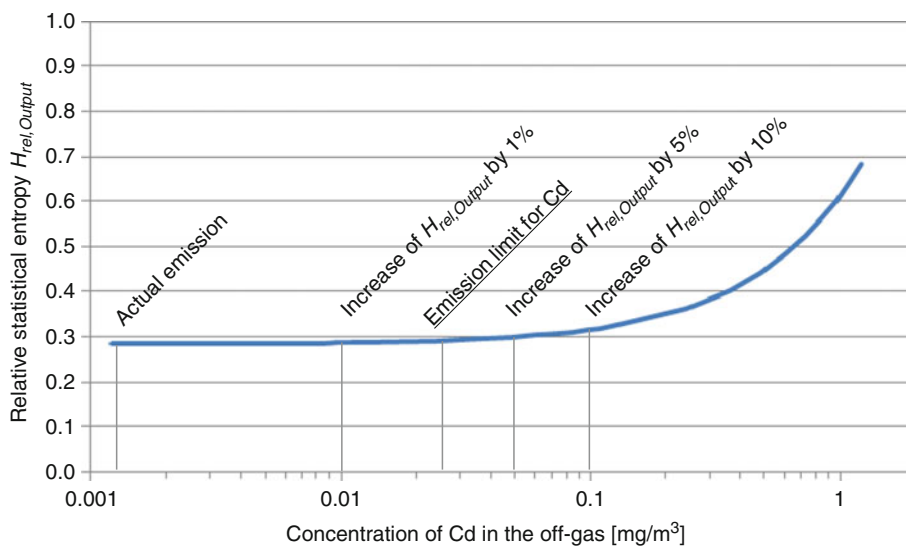
Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Figure 6
Occurrence of metals (%) in the bottom ash of a WTE facility (mass burn) after magnetic separation of iron [6]

determined correctly. The European Directive on Waste Incineration [5] provides the following emission limits: $\text{Cd} + \text{Tl} = 0.05 \text{ mg/m}^3$, $\text{Hg} = 0.05 \text{ mg/m}^3$, and $\text{Sb} + \text{As} + \text{Pb} + \text{Cr} + \text{Co} + \text{Cu} + \text{Mn} + \text{Ni} + \text{V} = 0.5 \text{ mg/m}^3$. For the sum of elements from Sb to V, Table 5 reveals

that an emission of 1.1 mg/m^3 would result in an entropy increase ($H_{rel,j,Output}$) of 1%. For entropy increases by 5% and 10%, the emissions would have to be 6.3 and 11 mg/m^3 , respectively. This calculation shows that, from an entropy point of view, the emission



Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Figure 7
Melting of bottom ash; copper flows. Some 80% of the copper in MSW could be recovered [9]



Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Figure 8
Illustration of the influence of the stack gas emission of a WTE facility on the relative statistical entropy ($H_{rel,Output}$). The emission limit for Cd is in the transition area between 1 and 5% increase in entropy

standard for this group of metals is rather stringent. In the case of mercury (Hg), the entropy approach would suggest an emission limit of 0.01 mg/m³. In this case, the actual emission standard of 0.05 mg/m³ seems rather high.

However, it should be noted that the transfer coefficients for metals in the stack gas have high uncertainties because they are the result of a single point measurement; also the geogenic background concentrations of metals vary by one to two orders

Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery. Table 5 Calculated emission values (mg/m³) as a result of entropy increase

	Increase of $H_{rel,j,Output}$		
	1%	5%	10%
As	0.004	0.01	0.02
Co	0.005	0.03	0.05
Cr	0.1	0.7	1
Cu	0.5	3.5	5
Mn	0.2	0.7	1.3
Ni	0.08	0.4	0.7
Pb	0.23	0.8	2.4
Sb	0.02	0.1	0.2
V	0.008	0.04	0.07
Total (As + Co + Cr + Cu + Mn + Ni + Pb + Sb + V)	1.1	6.3	10.7
Emission limit for As + Co + Cr + Cu + Mn + Ni + Pb + Sb + V [5]			0.5

of magnitude. In summary, it can be stated with high certainty that the actual emissions of metals are not entropy relevant and with some certainty that the present emission standards are in the transition area where emissions to the air would become relevant. Therefore it seems that there is a weak relationship between entropy generation and environmental impact. This means that the entropy approach cannot replace human and eco-toxicological research to fix emission limits but can be used as a first approximation tool where limits are missing or need to be assessed.

Future Directions

At present, research is carried out to extend the concept of SEA from chemical elements to compounds. For example, instead of examining the nitrogen balance of a wastewater treatment plant, one would consider the different compounds of nitrogen in the inflow and effluent of the plant, such as nitrite, ammonium, nitrous oxide, etc. The consideration of the chemical speciation of an element is important for substances like carbon and nitrogen. Since the speciation of an element is crucial for its eco-toxicity, it will be interesting to find out whether there is a relationship between entropy

and eco-toxicity. Another research topic is the testing of the applicability and usefulness of SEA in benchmarking studies for waste and wastewater treatment technologies and the employment of statistical entropy for the evaluation of resource consumption.

Bibliography

Primary Literature

1. Georgescu-Roegen N (1971) The entropy law and economic process. Harvard University Press, Cambridge, MA
2. Stumm W, Davis J (1974) Kann recycling die umweltbeeinträchtigung vermindern? In: Gottlieb Duttweiler-Institut für wirtschaftliche und soziale studien, Recycling: Lösung der Umweltkrise? Ex Libris, Zurich, pp 29–41
3. Rechberger H, Graedel TE (2002) The European copper cycle: statistical entropy analysis. *Ecol Econ* 42(1–2):59–72
4. Brunner PH, Morf LS, Rechberger H (2004) Thermal waste treatment – a necessary element for sustainable waste management. In: Twardowska I, Allen HE, Kettrup AAF, Lacy WJ (eds) Solid waste: assessment, monitoring and remediation. Elsevier, Kidlington, Oxford, pp 783–806
5. Directive 2000/76/EC of the European Parliament and of the Council of 4 December 2000 on the incineration of waste
6. Skutan S, Rechberger H (2007) Bestimmung von stoffbilanzen und transferkoeffizienten für die linie II der MVA Wels. Final Report, Institute for Water Quality, Resource and Waste Management, Vienna University of Technology, unpublished

7. Rechberger H (1999) Entwicklung einer methode zur bewertung von stoffbilanzen in der abfallwirtschaft. PhD-Thesis, Wiener Mitteilungen Band 158, Institute for Water Quality, Resource and Waste Management, Vienna University of Technology
8. Rechberger H, Brunner PH (2002) A new, entropy based method to support waste and resource management decisions. *Environ Sci Technol* 34(4):809–816
9. Mitterbauer G, Skutan S, Rechberger H (2009) Charakterisierung der rostasche der MVA Dürnröhr im Hinblick auf die rückgewinnung von metallen. Institute for Water Quality, Resource and Waste Management, Vienna University of Technology, unpublished
10. Zeltner C, Lichtensteiger T (2002) Thermal waste treatment and resource management – a petrologic approach to control the genesis of materials in smelting processes. *Environ Eng Policy* 3:75–86

Books and Reviews

- Brunner PH, Rechberger H (2004) Practical handbook of material flow analysis. Lewis Publishers, New York
- Kaufman S, Krishnan N, Known E, Castaldi M, Themelis N, Rechberger H (2008) Examination of the fate of carbon in waste management systems through statistical entropy and life cycle analysis. *Environ Sci Technol* 42(22):8558–8563
- Rechberger H (2001) An entropy based method to evaluate hazardous inorganic substance balances of waste treatment systems. *Waste Manage Res* 19:186–192
- Rechberger H (2001) The use of the statistical entropy to evaluate the utilisation of incinerator ashes for the production of cement. *Waste Manage Res* 19:262–268

Waste-to-Energy Ash Management in Europe

JUERGEN VEHLW

Karlsruhe Institute for Technology, Institute for Technical Chemistry, Karlsruhe, Germany

Article Outline

Glossary
 Definition of Subject
 Introduction
 Bottom Ashes
 Boiler and Filter Ashes
 Residues from Gas Cleaning
 Conclusions
 Future Directions
 Bibliography

Glossary

Boiler ash Coarse fly ash separated from the flue gas inside the boiler and directly discharged or discharged after boiler cleaning by soot blowing or rapping.

Bottom ash Solid residue discharged from the grate of a waste incinerator.

Dioxins (PCDD/F) A group of 75 congeners of polychlorinated dibenzo-p-dioxins (PCDD) and 135 congeners of polychlorinated dibenzofurans (PCDF); concentrations are given in toxic equivalent accounting for the different toxicity of the single congeners.

Filter ash Fine dust separated from the flue gas by filters, mainly by electrostatic precipitators.

Incineration Combustion of waste, in Europe preferentially in grate furnaces.

Leaching Elution of inorganic (especially heavy metal) or organic species out of solid materials into liquids; leaching standards are typically regulating the access to disposal or the utilization of a material; leaching is measured by standardized tests.

Scrubbing Chemical gas cleaning step using wet or dry neutralizing agents; wet systems are typically designed in two stages.

Definition of Subject

In Europe, the environmentally compatible management of solid residues from waste incineration became a major topic of concern in the first decade of the new millennium. Tightened access criteria to landfills promoted in a number of countries, especially in the European Union, aim for beneficial use of the bottom ash. Furthermore, several attempts are seen to stabilize respectively inertize fly ash and gas cleaning residues so they need not be disposed of in hazardous landfills. Organic content and leaching stability are the most important environment-related quality parameters to be addressed.

Introduction

Some time ago, the debate about thermal processes was mainly driven by potential risks of air emissions, especially those related to dioxins. Meanwhile, the gas

cleaning devices implemented in municipal solid waste incineration (MSWI) or, as they are called in some areas, waste-to-energy (WTE) plants are among the most effective ones found in any technical process, and the interest is nowadays more focused on the quality of the solid residues from those processes. The actual objectives are to produce as much as possible inert bottom ashes and to enable their utilization as secondary building materials. This is especially promoted in the Netherlands, Denmark, and Germany, recently also in France. A further focus is the inertization and safe disposal – or even utilization – of the filter and boiler ashes as well as of the residues collected in the Air Pollution Control (APC) systems of WTE plants.

Already in early times attempts are seen to utilize bottom ashes from waste incineration plants. In the Frederiksberg incinerator, built in 1903, bricks were produced from bottom ash for utilization in the building sector [1]. However, the calorific value of the waste at that time and the resulting combustion temperature were most likely to generate a bottom ash quality which was far beyond the standards required and achieved today. The described procedure was only maintained for few years, supposedly due to market constraints based on insufficient quality.

After World War I a tendency was seen not to burn but to melt the waste, again to produce secondary building materials. This practice, too, was finally not successful on the market because of problems with the refractory liners of the furnaces, varying product quality, and especially the high costs of the processes [2].

Starting the last decade of the twentieth century all kinds of secondary treatment processes have been developed to tailor the residue quality according to special needs. Secondary measures, however, are expensive and hence the better approach is an optimized control of the combustion process to

- Guarantee an excellent burnout of carbon compounds
- Promote the volatilization of heavy metals like Hg and Cd out of the fuel bed
- Fixate lithophilic elements in the silicatic and oxidic matrix of the bottom ash, thus reducing their leachability

The following discussion of the quality of residues from modern waste incineration plants will follow these objectives. Finally, this entry will investigate which rational options exist to inertize and eventually utilize bottom, boiler, and filter ashes, and flue gas cleaning residues. All considerations are based mainly on the results of an international perspective on the characterization and management of waste incineration residues, published by the International Ash Working Group in 1997 [3].

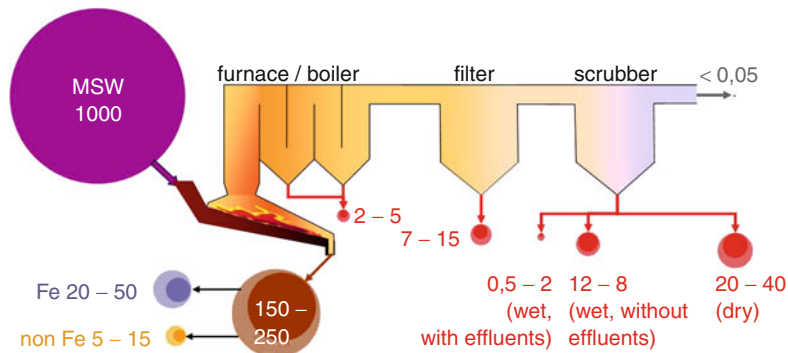
Mass Streams in a WTE Plant

The basis of all discussions about waste incineration residues is the knowledge of the different mass streams in a municipal solid waste incinerator. Figure 1 shows average ranges for these streams as found in modern mass burning systems. The air consumption is approx. 4,500 m³/Mg of waste.

State-of-the-art plants in most industrialized countries produce typically between 150 and 250 kg of bottom ash per 1 Mg of burnt waste. The respective numbers in Japan are in the order of 100–120 kg due to different waste collection and eventual pre-treatment strategies.

Most published numbers include the grate siftings which are in many countries kept separate from the bottom ash. The mass flow of siftings depends on the type of grate and its time of operation. It accounts typically for less than 5 kg/Mg of waste. The siftings may increase the amount of unburnt matter in the bottom ash. In view of utilization, however, the inventory of metallic Al eventually dripping through the grate voids is of much higher concern. Metallic Al tends to form hydroxides in contact with water and this reaction is accompanied by a swelling of the material.

The production of boiler ash depends on the type of boiler and on the amount of dust released from the grate and carried over in the flue gas. Boiler ashes should be treated together with the filter ashes due to their similar level of toxic heavy metals and organic compounds. In some countries, this has already been enforced by legislative regulations. The data presented for filter ashes reflect the situation in modern plants that try to establish so-called gentle combustion with dust loads down to less than 2 g/m³ [4].



Waste-to-Energy Ash Management in Europe. Figure 1
Mass balance of a municipal solid waste incineration plant (in kg/Mg waste)

The mass flow of APC residues shows actually the highest variation of all residues. Wet scrubbing systems with liquid effluents discharge produce 0.5–2 kg of dry neutral sludge per Mg of burnt waste. In wet systems with internal or external evaporation of the scrubbing solutions additional 11–16 kg of neutralization products increase the total solid residues to approx. 12–18 kg/Mg. Whereas wet scrubbing operates close to stoichiometry and hence at minimum residues, in case of semi-dry or dry systems the necessary surplus of unreacted neutralizing reagents increases the total amount of scrubbing residues to 20–40 kg/Mg of waste.

Bottom Ashes

Disposal and Utilization Regulations

Waste incineration is performed to produce an inert residue, the bottom ash, which meets the respective disposal standards. Many countries aim for utilization of this residue stream in order to save space on landfill sites.

The disposal of residues is in the EU regulated by the EU landfill directive 1999/31/EC [5] and the Council Decision 2003/33/EC [6] which contains the landfill access criteria. For utilization of residues, however, no common regulations exist and the various countries have their own protocols.

Table 1 compiles selected standards laid down in the German Waste Storage Ordinance [7] for disposal on landfill class 1 – which are identical with

Waste-to-Energy Ash Management in Europe. Table 1
Selected German standards for disposal and utilization of WTE bottom ash

	Unit	German landfill class 1	Road construction (LAGA)
LOI	wt%	3	
TOC	wt%	1	1
DIN EN 12457-4			
Cl	mg/L	1,500	250
Cu	mg/L	1	0.3
Zn	mg/L	2	0.3
Cd	mg/L	0.05	0.005
Hg	mg/L	0.005	0.001
Pb	mg/L	0.2	0.05

those in the Council Decision – and for utilization of bottom ash in road construction which are laid down in a memorandum of the LAGA, a German working group of representatives of the ministers of environment of the federal states [8]. The leaching standards have to be measured according to the EU leaching protocol [9] which is in the table indicated by the respective national test DIN EN 12457-4. The table shows only small differences between the requirements for disposal and utilization. The challenge is to reach the utilization quality without further post-combustion treatment.

In the case of utilization as secondary building material, additional standards are set for mechanical properties like density, mechanical strength, grain size distribution, or freeze-thaw-stability. This aspect will not be discussed here.

Characterization

Burnout and Organic Ingredients The “burnout” is the key parameter for disposal as well as for utilization of bottom ashes. The German Waste Storage Ordinance sets a total organic carbon (TOC) limit of 1 wt.-% for disposal on a class I landfill and the same number is found in the LAGA memorandum for utilization in road construction.

In modern well-operated WTE plants, the TOC in bottom ashes is typically well below 1 wt.-% [10, 11]. Special combustion trials in the TAMARA test incinerator in Karlsruhe (Germany) demonstrated that an increasing heating value of the feed, and the resulting higher bed temperatures, improve the burnout of bottom ash (see Fig. 2) [12].

The TOC of bottom ashes consists mostly of elementary carbon; some organic compounds are also found that cover the spectrum from short-chain compounds [13] up to low volatile species such as PAH or PCDD/F. Typical concentrations of organic

compounds in the various solid residues are compiled in Table 2.

The data are taken from modern facilities [12–15]. The PCDD/F numbers are given in terms of international toxic equivalence data (I-TE). It is evident that the organic pollution is higher in the boiler and fly ashes than it is in the bottom ash.

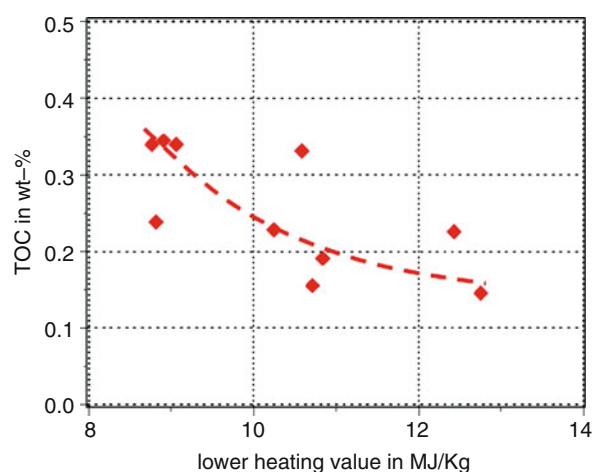
It should be noted that the I-TE levels detected in the bottom ashes of modern incineration plants were in the same order of magnitude as found in uncontaminated soils in Central Europe [16].

Chemical and Mineralogical Characterization

The mass and volume reduction occurring during combustion of solid wastes causes an enrichment of a number of heavy metals in the bottom ashes compared to their concentration in the earth's crust. This is demonstrated by the concentration ranges of selected species depicted in Fig. 3. The data are compiled from numerous publications and from the Handbook of Chemistry and Physics.

Some heavy metals, for example, As, Cd, Hg, and Pb are to a great extent volatilized out of the fuel bed, however, also these metals show in most cases higher concentrations in the bottom ashes than in the lithosphere.

In addition to the chemical analysis, a geochemical and mineralogical characterization provides useful information in view of the long-term behavior of a material. Bottom ashes can be characterized as a mixture of silicate and oxide phases. Some typical



Waste-to-Energy Ash Management in Europe. Figure 2 Residual carbon in TAMARA bottom ashes versus heating value of the feed

Waste-to-Energy Ash Management in Europe. Table 2 Concentration ranges of organic compounds in bottom, boiler, and filter ashes (PCDD/F in I-TE)

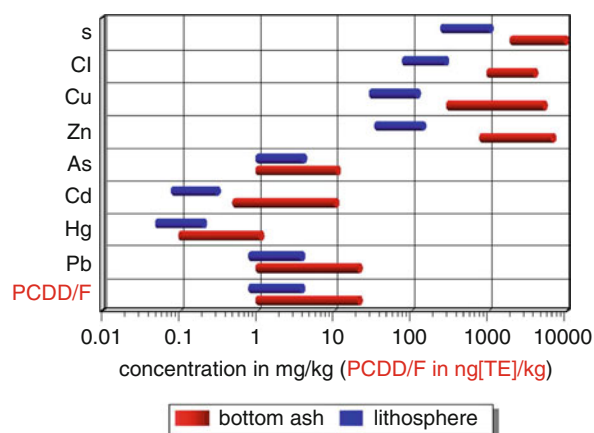
	Bottom ash (ng/g)	Boiler ash (ng/g)	Filter ash (ng/g)
PCDD/F	<0.001–0.03	0.02–1	0.1–10
PCB	<5–50	4–50	10–250
PCBz	< 2–20	200–1,000	100–4,000
PCPh	<2–50	20–500	50–10,000
PAH	<5–10	10–300	50–2,000

mineral phases found in these residues are shown in the micrographs in Fig. 4 [17].

Some of these phases can provide information about the temperature the material has been exposed to on the grate. This important parameter controls mainly the fate of elements in the combustion chamber, but is not well known in full scale plants. The knowledge of formation temperatures of single phases and the specific search for high-temperature phases are promising ways to obtain better information in this area [18].

Post-combustion Treatment

As mentioned earlier, a number of countries have or are planning to set standards for the utilization of bottom ash. The major application area is road construction



Waste-to-Energy Ash Management in Europe. Figure 3 Concentration ranges of selected elements and PCDD/F in bottom ash and in the earth's crust

where ash is used in the support layers, mainly under water-tight capping. Among the few countries which have more or less country-wide common regulations is Germany which will be discussed more in detail below. Similar practice is followed in the Netherlands, Belgium, and Denmark. Countries like Sweden decide on utilization case by case.

The German guideline regulating use of ash in road construction is the LAGA memorandum mentioned earlier [8]. It requires mechanical pre-treatment, scrap removal, and aging of the ashes for 12 weeks.

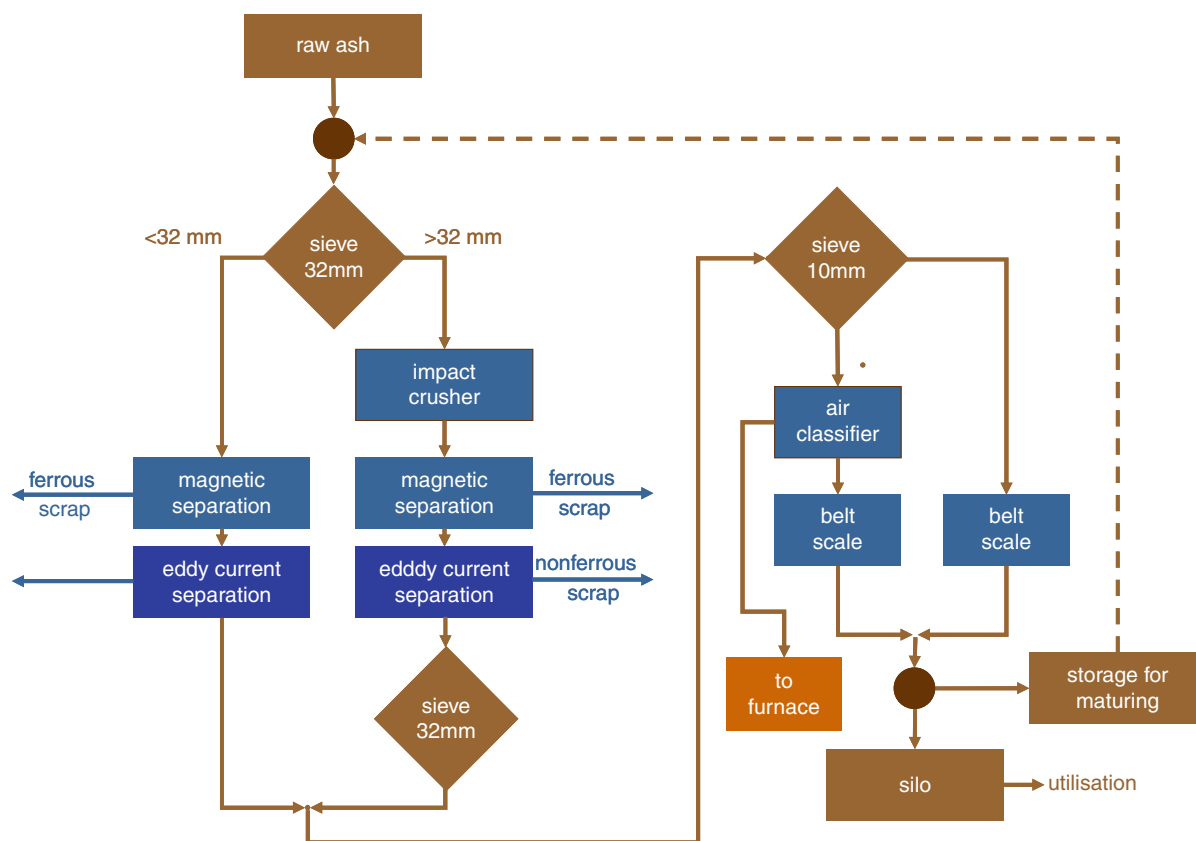
The mass flow scheme in Fig. 1 indicated already that raw bottom ashes contain significant amounts of ferrous and non-ferrous metal scrap. It is common practice that at least the ferrous, but nowadays in most plants also the non-ferrous, metals are recovered if the final destination of the bottom ash is a landfill. Whereas revenue from ferrous scrap may often only cover the recovery costs, non-ferrous metals, mainly Al and Cu, have a high market value which cares for good economy.

Hence, scrap removal is the only pre-treatment if bottom ashes are landfilled. If utilization is envisaged, the obligatory post-combustion treatment is much more extended. A typical flow sheet of advanced ash treatment as performed at the Hamburg waste incinerator MVR is shown in Fig. 5.

The procedure starts typically already in the quench tank of the incinerator where a surplus of water reduces the salt content [10, 20]. After discharge, the ashes are stored for few days for de-watering before they undergo further treatment which consists of a sieving to remove bulky fractions and magnetic separation of ferrous



Waste-to-Energy Ash Management in Europe. Figure 4 Micrographs of minerals in bottom ashes: glass formed during combustion (left), gehlenite $\text{Ca}_2\text{AlAlSiO}_7$ in glassy matrix (center), magnetite, Fe_3O_4 (right)



Waste-to-Energy Ash Management in Europe. Figure 5

Grate ash pre-treatment at the Hamburg waste incineration plant MVR (Adopted from [19])

scrap which is followed by an eddy-current separation of non-ferrous metals.

The metal fractions are sold to recycling companies. Considering the total amount of bottom ashes of 4–5 million Mg in Germany and a recovery efficiency of 60–70% more than 300,000 Mg of iron scrap and with a recovery efficiency of 30–40% up to 50,000 Mg of non-ferrous metals can be recovered. Especially, the latter fraction accounts for a not unwanted revenue for the operator.

The ashes are then stored for aging or maturing for 12 weeks and after the respective compliance tests, the ashes can be utilized as secondary building materials.

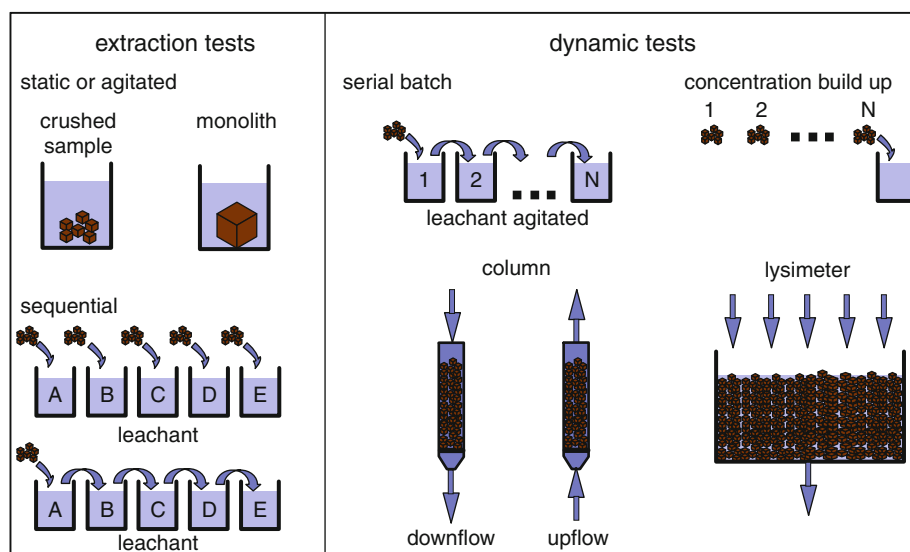
Leaching Properties

Leaching Fundamentals The chemical composition of a product may not allow determining its environmental impact. This is far more depending on the

leaching stability of the material in question. Even if the matrix and the speciation of single elements are known, a reliable theoretical prediction of the short- and long-term behavior is almost impossible. The most important parameters influencing the leaching stability of a material are listed below:

- Its chemical composition
- Its chemical/geochemical/mineralogical speciation
- The fraction of a species available for leaching
- The particle morphology
- The properties of the leachant, especially its pH or the presence of complexing constituents
- The liquid-solid ratio (LS) in the leaching system

It seems evident, that no single or simple test procedure will deliver results that allow a sound evaluation of impacts on the environment. In fact a great number of different tests have been developed to get detailed



Waste-to-Energy Ash Management in Europe. Figure 6
Principles of leaching tests

information about the leaching properties of residues from waste incineration.

There are two principal categories of test procedures: extraction tests and dynamic tests. Indicator tests are in most cases extraction tests. The most common principles of these test categories are comprised in Fig. 6.

Extraction tests allow the determination of leaching equilibria. If the leachate is analyzed in time increments before equilibrium is reached, information on the kinetics of the system can be obtained. An example of such tests is the Dutch tank leaching test [21] for stabilized materials. This test also gives indication of the major parameters controlling the leaching process, for example, diffusion or solubility. Sequential tests in different leachants of increasing chemical strength are often used to investigate the chemical bond of specific elements in the matrix.

Dynamic tests are applied to reveal the leaching kinetics of granular material. One of these is the Dutch column leaching test [22]. This test is typically performed up to an accumulated liquid/solid ratio of 10 and enables the modeling of contaminant release over a period of 50–100 years.

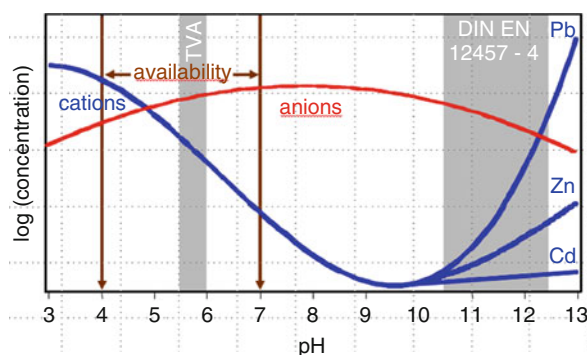
In practice, a material in question is subject to a number of different tests that are selected to model as close as possible the envisaged disposal or utilization

scenario. After the fundamental properties have been acquired, an indicator or compliance test – in most cases a standardized test procedure – is chosen to assess the quality of an actual sample in short time and with limited effort [23]. Since the early 2000s, the European Standardization Committee CEN has released unified test methods which will in future be applied in all EU countries [9].

The most important parameter influencing the results of a leaching test is the pH of the leachate [3, 24]. Figure 7 shows a schematic overview of the influence of the pH of the leaching solution on the solubility of metal cations and anions in aqueous solutions. It is well known that most heavy metals show rather low solubility in the weak alkaline range. Their solubility increases with decreasing pH. In the alkaline region, different metals behave differently: some (e.g., Cd) stay insoluble with increasing pH. Others, the amphoteric ones, are more or less solubilized if the pH is elevated. The amphoteric metal of highest interest in waste incineration is Pb.

Metals which tend to form anions in aquatic solutions like V, Cr, or Mo, have their highest solubility close to the neutral point.

The graph in Fig. 7 indicates the ranges of pH used in selected leaching tests of bottom ashes. In the DIN EN 12457-4 test with a LS ratio of 10, a duration of



Waste-to-Energy Ash Management in Europe. Figure 7 Influence of the pH and the solubility of metal cations and anions

24 h, and no pH control, the final pH values can vary between 10.5 and 12.5. These varying conditions have severe impacts especially on the test results for Pb as can be deduced from Fig. 7.

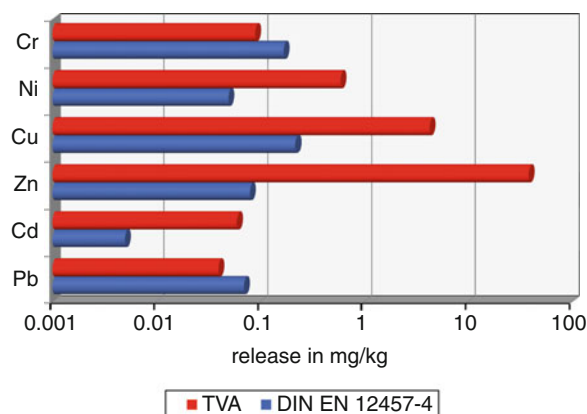
The Swiss TVA test [25], a compliance test with two steps, 24 h each, LS = 10, is characterized by a rather constant pH of 5.5–6 due to gaseous CO₂ injected permanently into the test solution.

Constant pH values are used for the Dutch availability test [22] which provides information about the leaching potential under assumed “worst” environmental conditions. The cation solubility is tested at a pH of 4 and that of anions at a pH of 7. The sample has to be finely ground in order to exclude any inhibition of the leaching by diffusion and the liquid-solid ratio is kept at 100 to avoid saturation effects in the solution.

Due to the different test conditions various leaching tests will produce different results when run on one and the same material. An example is shown in Fig. 8, where one bottom ash sample was tested using the Swiss and the German test procedure [26].

The bar plot indicates that Ni, Cu, and especially Zn show a significantly higher elution in the TVA than in the DIN EN test due to the formation of hydrogen-carbonates.

Effect of Aging and Utilization Potential In order to optimize the total burnout, the combustion temperature, and thus, the fuel bed temperature have been elevated in WTE plants compared to the operation

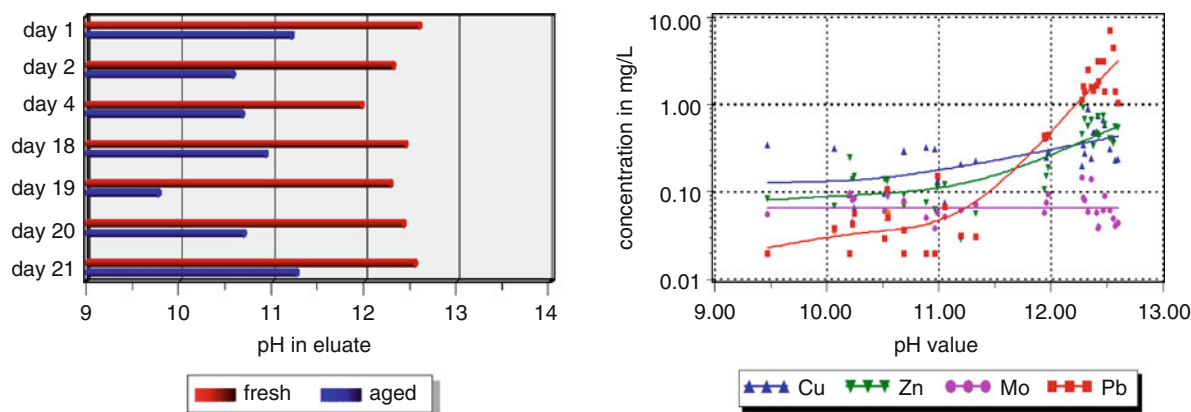


Waste-to-Energy Ash Management in Europe. Figure 8 Leaching results of the Swiss TVA and the German DIN EN 12457-4 test on one bottom ash sample (Adopted from [26])

mode during the 1980s. As an effect of such operation changes, a higher formation of CaO can be seen. The pH value of fresh bottom ashes is often exceeding 12. According to the German LAGA memorandum [8] bottom ashes have to be stored for 12 weeks prior to utilization in road construction for aging or maturing. During this time the uptake of CO₂ from the air converts the earth-alkali oxides into carbonates and neutralizes part of the alkalinity. Hence, aged bottom ashes establish a pH of about 10–11 in the DEV S4 test.

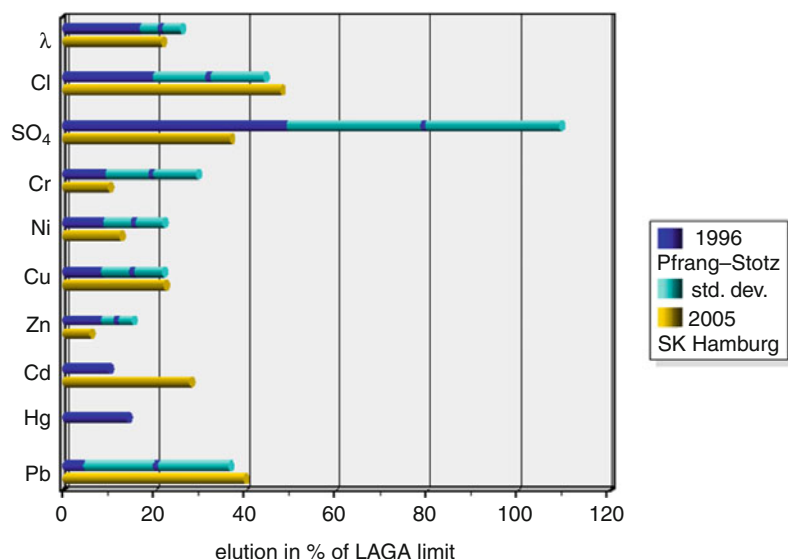
Data from a test program in a German full scale waste incineration plant illustrate the effect that aging has on the pH of bottom ash and on the leaching of metals [11]. The bar plot in the left part of Fig. 9 documents that the pH of fresh bottom ash in the DIN EN 12457-4 test exceeds in almost all test trials 12 and drops down by about two units during the aging process. As shown in the right graph of Fig. 9, this pH change has no effect on the leaching properties of Mo which is present mainly as molybdate. The leaching stability of Cu and Zn is moderately improved in the aged material whereas the leaching of Pb is reduced by almost two orders of magnitude.

This strong interdependency is responsible for the strange situation that due to German regulations fresh bottom ashes from some plants do not comply with the landfill standards while after aging they are excellent secondary building materials.



Waste-to-Energy Ash Management in Europe. Figure 9

pH value of fresh and aged bottom ash samples (left) and DIN EN 12457-4 test results on these samples for four metals (Adopted from [11])



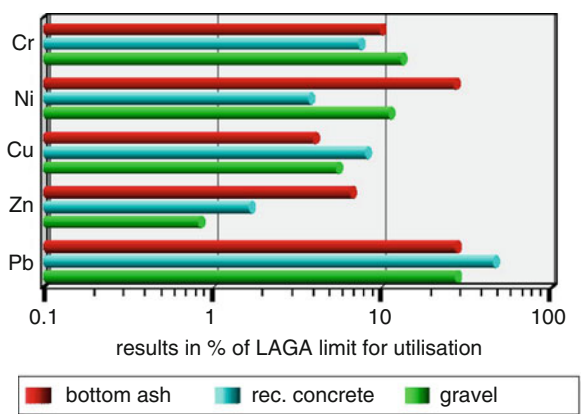
Waste-to-Energy Ash Management in Europe. Figure 10

Bottom ash leaching test results obtained by the German compliance test in 1996 (Adopted from [27]) and in 2005 (Adopted from [19]) (λ: electric conductivity)

Figure 10 shows the results for 26 samples taken routinely in 1994 over 1 year from an industrial ash treatment site [27]. The test at that time was almost identical with the DIN EN 12457-4 procedure. The results for the environmentally interesting heavy metals were always well below the respective standards. The only component exceeding the limit in a few cases was sulfate. This limit has been set to protect concrete structures from corrosion attack.

In 2005, equivalent data were published for bottom ashes of the Hamburg, Germany, waste incineration plants [19]. Except for the much better sulfate values the results are almost the same (Fig. 10).

Meeting standards is a prerequisite to get approval for utilization, more detailed information about the potential environmental impact may be achieved by running the same tests on bottom ashes and other building materials. This has been done using aged



Waste-to-Energy Ash Management in Europe. Figure 11 Leaching test results obtained by the German compliance test on bottom ashes, recycled concrete, and gravel (Adopted from [28])

bottom ashes from a German waste incinerator, secondary building material recycled from a demolished highway bridge, and gravel from the river Rhine. The results are shown in Fig. 11 [28]. The bar plot documents that the difference in leaching stability of the three materials is surprisingly small, although there is no doubt, that the inventory of the selected metals in bottom ash is much higher than in concrete and especially in gravel.

Utilization of Bottom Ashes in Europe

According to the above discussed bottom ash quality, it can be stated that bottom ashes from modern and well-operated WTE plants meet the LAGA limits for utilization easily and do not behave significantly differently from conventional building materials. Hence, there is no reason not to utilize – after careful testing – such materials for e.g., road construction [29].

This use is common practice in countries with geological conditions hampering the siting of landfills like the Netherlands, Belgium, or Denmark. These countries utilize 90% and more of the bottom ashes. The respective number for Germany is approx. 85%. Some other countries like France are nowadays also encouraging bottom ash utilization.

A different strategy is followed by the Swiss authorities. According to their regulations, bottom

ashes are categorized as reactive residues. Only stone-like materials are accepted as building materials and stone-like refers to the concentration and not to the mobility of a single constituent. Since the bottom ashes contain higher amounts of heavy metals than the lithosphere (see Fig. 3), almost no utilization is practiced. Bottom ashes have to be disposed of or they must be converted into stone-like materials by adequate measures.

The practiced management strategies for bottom ashes from European WTE plants are shown in Fig. 12. The data are taken from Eurostat, the statistical office of the EU [30], and from CEWEP, the Confederation of European Waste-to-Energy Plants [31].

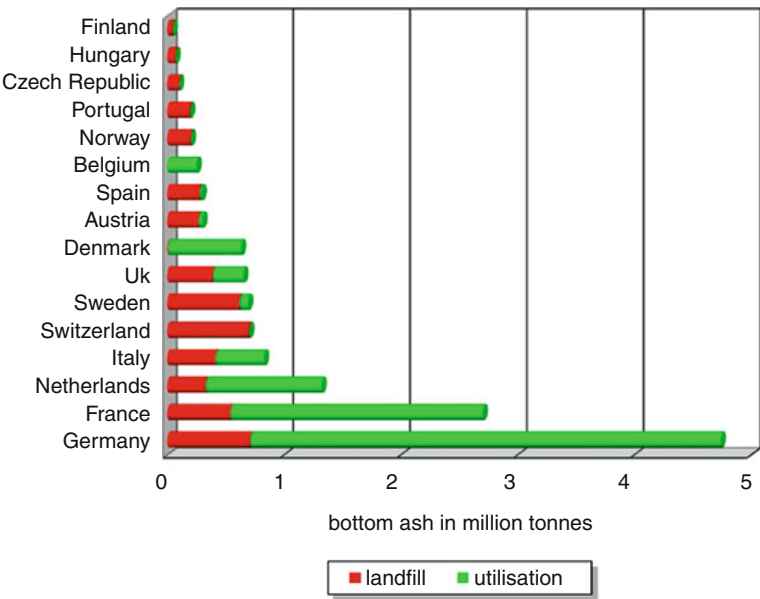
As can be seen, most countries with a high share of waste incineration and hence high bottom ash figures utilize a significant fraction of that material, mainly in road construction.

Quality Improvement by Sintering

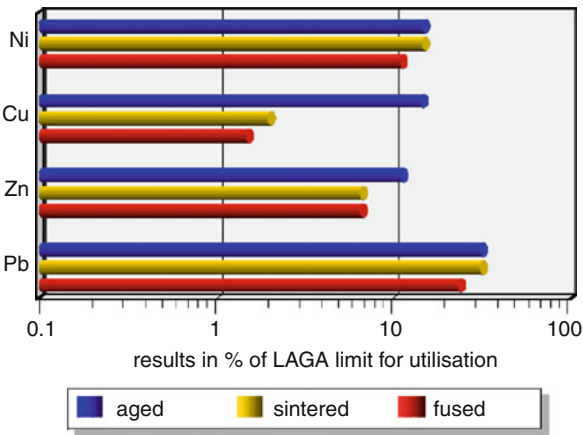
The above presented high leaching stability of bottom ash needs to be achieved permanently. This leads to the question, how to guarantee such high quality. A self-evident approach is the establishment of high temperatures in the fuel bed to volatilize mobile metals and halogens out of the bed material and to immobilize lithophilic species by sintering. Since sintering is a solid phase respeciation with stabilization of many heavy metals in the silicate-rich lattice of the bottom ash, a higher residence time of the ash at higher temperature in the furnace promotes the formation of inert species.

This strategy has been investigated in laboratory scale by sinter experiments using fresh bottom ashes from two German incineration plants [10]. The ashes were treated thermally under atmospheric air at temperatures of 850°C, 1,000°C, and 1,300°C for 30 min each. At the latter temperature, the material was melted. The resulting DEV S4 leaching data of the products of these tests are presented in Fig. 13.

The graph indicates a significant improvement of the leaching stability of four selected metals by the treatment. At 850°C and 1,000°C, comparable effects were observed. The fusion, however, did not improve the elution stability significantly. This finding is supported by the comparison of test results from bottom ashes with those published for molten residues



Waste-to-Energy Ash Management in Europe. Figure 12
Production, disposal on landfills, and utilization of bottom ashes in all European countries which operate WTE plants in 2008



Waste-to-Energy Ash Management in Europe. Figure 13
Leaching test results of thermally treated bottom ashes and concrete

from high-temperature processes like Thermoselect or the Siemens Thermal Waste Recycling Process [32].
The stabilization by sintering was also validated in semi-technical experiments with fresh bottom ashes from a full scale incineration plant [33].

Based on these results, it can be concluded that sintering at temperatures of 850°C has a stabilizing effect upon heavy metals. The energy consuming – and thus more costly – fusion, however, does not pay since no significant further fixation could be observed. Hence, a simple in-plant measure to produce bottom ashes of high leaching stability can be recommended: The bed material should be kept at high temperature for as long as possible at the back end of the grate.
This is common practice in Europe today, attempts to implement melting or fusion processes into conventional waste incineration plants or to establish a market for combined processes with molten slag like Thermoselect [35], the Siemens Thermal Waste Recycling Process [36], or the Von Roll RCP Process [37] failed in Europe – in spite of strong efforts during the 1990s. Main reasons were energy consumption and complexity of the processes which caused high process costs as well as technical problems.
Other stabilization processes like stabilization by the addition of phosphates (WES-PHIX process) are also not common since the quality of bottom ashes from

well-operated incineration plants meets all requirements for safe long-term disposal or even utilization – as has been laid out above.

Boiler and Filter Ashes

Characterization

Boiler ashes are deposited at temperatures between approx. 800°C and 200°C, filter ashes typically at temperatures slightly below 200°C. That is why heavy metals and other species which are volatilized inside the combustion chamber are to some extent precipitated on the dust surfaces. The concentration of volatile elements like Zn, As, Cd, or Pb can exceed that in grate ashes significantly as is shown in Fig. 14. The inventory of halogenated organic micro-pollutants like PCDD/F or PCB which are mainly synthesized inside the boiler or of PAH which are an indicator for incomplete combustion is also increased compared to that in bottom ashes (compare Table 2).

Since filter ashes are precipitated at lower temperatures than boiler ashes and since they have a much finer particle size distribution and offer higher surfaces for

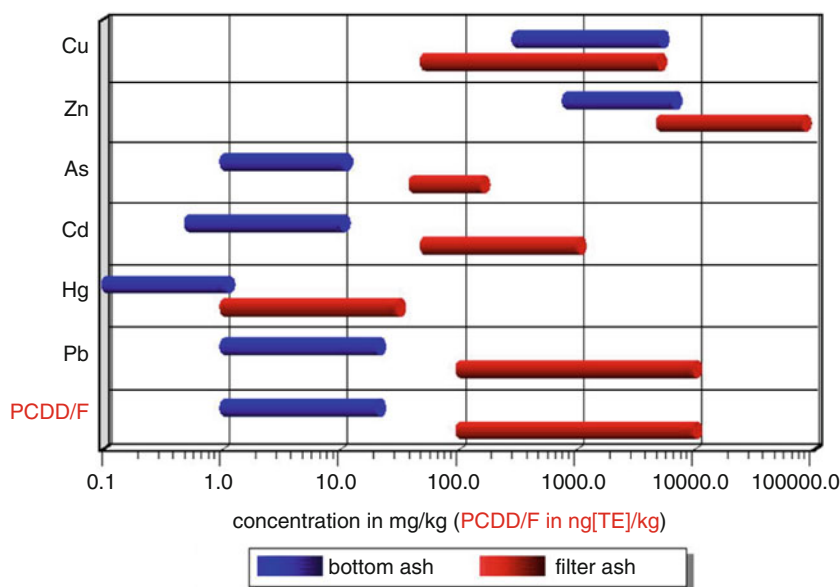
condensation their concentrations of heavy metals, of halogenides, and of organic micro-pollutants are typically higher.

Disposal Practice

Due to their elevated pollutant inventory boiler and filter ashes have to be characterized as hazardous waste and must not be combined with the bottom ashes. They have to be typically disposed of on special – and expensive – disposal sites, preferentially in the underground. A compilation of filter ash production as well as disposal strategy in European countries is found in Table 3. The data have been compiled from CEWEP member's questionnaires [31].

The table documents that the by far prevailing management strategy is the disposal on hazardous landfill sites.

In Germany boiler and filter ashes can be used in old salt mines for backfilling of caverns. This practice is according to national regulations acknowledged as utilization. The same strategy is followed in Norway where filter ashes can be “utilized” for backfilling an old granite quarry on the small island Langøya. Norwegian



Waste-to-Energy Ash Management in Europe. Figure 14

Concentration ranges of selected elements and PCDD/F in bottom and filter ash

Waste-to-Energy Ash Management in Europe. Table 3 Filter ash production and management in selected countries [CEWEP 2006] (¹estimated)

Country	Year	Production (Mg)	Management strategy
Austria	2004	40,000	Mainly salt mine backfilling, hazardous landfill
Belgium	2004	50,000	Solidification and hazardous landfill
Czech Republic	2004	6,000	5,000 hazardous landfill, 1,000 underground mine
Denmark	2004	86,834	100% recycled (salt mine after special treatment)
France		400,000	
Germany	2005	650,000	Utilization in salt mine, minor amount hazardous landfill
Hungary	2005	13,585	Hazardous landfill
Italy	2004	71,814	
Netherlands	2003	82,200	50% utilized in mines and construction (e.g., asphalt filler)
Portugal	2005	40,000	Inertization before hazardous landfill
Spain	2005	94,420	Hazardous landfill
Sweden	2004	138,000	Landfilling + utilization in Norway (Langøya)
United Kingdom ¹	2000	78,125	Hazardous landfill

data were not available but some filter ashes from Sweden are going to that place as well.

A different utilization of filter ashes is practiced in the Netherlands. Out of the noted 50% utilized ashes a certain amount is used as filler in asphaltic layers on roads, others are shipped to German salt mines for backfilling.

Stabilization Processes

The high expenses for a sustainable final disposal of boiler and filter ashes were reason for numerous attempts to detoxify these materials in order to get access to less expensive disposal routes or even to utilization scenarios. A broad spectrum of different processes has been proposed and tested in different scales. Table 4 tries to categorize the various treatment options.

It is evident that solidification or stabilization processes do not alter the toxic inventory of a material. The inertization by chemical reactions or the formation of a diffusion barrier does only last for a limited time. Two processes have been in full scale application for many years: the “Bamberg Model,” where filter ashes are stabilized on a landfill by mixing with the neutralization sludge of the wet scrubber effluents [39], and in Switzerland the stabilization of filter ashes by cement after washing [40].

Waste-to-Energy Ash Management in Europe. Table 4 Procedures for treatment of boiler and filter ashes

Principle	Process
Solidification/stabilization	Without additives (Bamberg Model) Cement based systems (e.g., Portland cement, alinite) Waste pozzolanic systems (e.g., coal fly ash) Chemical stabilization (e.g., sulphides, TMT 15™) Organic additives or matrix (e.g., bitumen)
Thermal treatment	PCDD/F destruction (Hagenmaier drum) Sintering (mineral respeciation) Fusion (melting without additives) Vitrification (melting with additives)
Combined processes	Acid extraction + sintering (3R process) Stabilisation with FeSO ₄ , oxidation, sintering (Ferrox process)

Thermal treatment can be performed at moderate temperatures in the so-called Hagenmaier Drum at 400°C to destroy dioxins [41]. This process is actually used in Japan; in Germany only few plants operated it in the 1990s as pilot plants.

Another option is vitrification at high temperatures ($>1,300^{\circ}\text{C}$) to produce glassy products. Vitrification is mainly favored in Japan as is the case for the bottom ashes. The molten products are distinguished by excellent leaching stability. Care has to be taken to avoid air pollution by evaporation of metal compounds. The energy consumption of all of these processes, however, is very high and is by far not compensated by the potential to recover metals. That is why such processes – although developed as described for the bottom ashes – did not conquer the market in Europe.

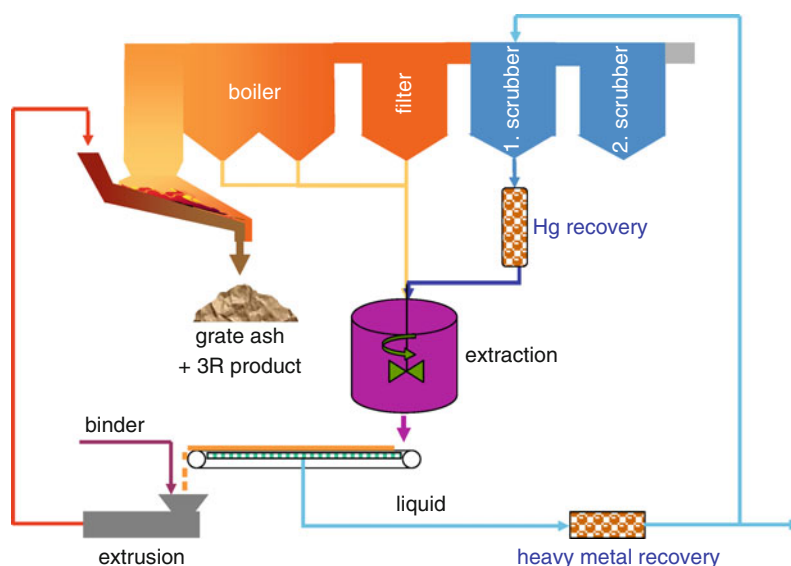
A third strategy – more in line with the demand for simple and in-plant measures – is followed by the 3R process which combines an acid extraction of soluble heavy metal compounds (by use of the acid flue gas cleaning solution) with a thermal treatment of the compacted extraction residues in the combustion chamber [42]. A plant equipped with this process has no filter and boiler ash stream and with that no need for a separate disposal of these critical residues. A scheme of the process is shown in Fig. 15. The technical demonstration revealed that the grate ashes from a 3R process which contain the leached filter ashes show excellent and not negatively altered leaching properties of heavy metals and that the PCDD/F are almost totally destroyed [43].

A similar process called FLUWA is practiced in a number of Swiss waste incineration plants. A main target in the further development of this process is the recovery of heavy metals out of the extraction solution [44].

The FerroX process for treatment of filter ashes and APC residues involves washing of the residues in a ferrous sulfate solution and contemporary oxidation of the iron to form insoluble iron hydroxides and oxide hydrates [45]. The products have the quality to be landfilled or they can for quality improvement be fed back into the combustion chamber for further thermal treatment. Tests in a pilot plant documented the almost total destruction of PCDD/F and an excellent leaching stability of heavy metals after such sintering [46, 47].

Residues from Gas Cleaning

The residues from wet gas cleaning without water discharge and those from dry or semi-dry APC systems carry high levels of soluble salts, especially of alkali and earth-alkali chlorides or sulfates. Due to the high solubility, a safe disposal can only be guaranteed on special and expensive disposal sites. Attempts have been made to utilize parts of the ingredients of these residues in order to minimize the disposal problem. The challenge



Waste-to-Energy Ash Management in Europe. Figure 15
Scheme of the 3R process

is the closing of the chlorine cycle. Different processes to recover NaCl [48], HCl [49], Cl₂ [50], or gypsum [49] have been tested. All such processes can only be successful if they end up with high quality products and if there is a long-term market for the products. Today, for example, in Germany only few MSWI plants produce HCl [51]. All other processes did not make it to the market due to unfavorable economy.

As mentioned above already, in Germany the disposal of gas cleaning residues and filter ashes for backfilling of caverns in old salt mines is acknowledged as utilization. This practice is widely used, to some extent also by neighboring countries.

Conclusions

European waste incineration plants aim since the early 1990s for the production of high quality bottom ashes for their safe disposal or even utilization as secondary building material. In line with this development also the sustainable management of boiler and filter ashes as well as of gas cleaning residues became of central concern.

Recommendations in that direction have been elaborated in the 1980s and 1990s; respective processes and operation protocols have been developed and have meanwhile been widely adopted by modern plants all over Europe. The major principles for achieving good management of WTE residues are compiled below.

- Adequate combustion control and careful sintering of the bed material at the back end of the grate guarantee an excellent burnout and cause a good fixation of heavy metals.
- Simple washing of the bottom ashes, preferentially in-plant in a modified quench tank, reduces the leaching of chlorides to very low levels.
- The resulting products have a high potential for utilization, for example, according to German regulations in road construction.
- Post-combustion treatment of bottom ashes increases the incineration cost without improving the elution stability significantly.

Most problems in the field of residue management are well understood today and in most cases appropriate technologies exist already. However, further development, especially in view of optimized energy

recovery, is expected. It is obvious that primary and in-plant measures have to be preferred rather than secondary post-combustion techniques.

Future Directions

The process of unification of waste management in the EU member states will continue, and it is most likely that in future all countries will enact a landfill ban for reactive waste. It can also be envisaged that in all member states recycling will catch up with the actually leading countries. This development in line with the expected convergence of economy and lifestyle will promote WTE in the EU. The quality of residual waste will not change dramatically in view of its calorific value, but their inventory of some ingredients, especially of metallic species, will hopefully decrease by advanced recycling strategies.

With the increased application of WTE the generation of solid residues from this process will increase as well. The technological progress should care for a higher quality of bottom ashes. However, whether this will go along with an increased utilization has to be waited for. The niche market as secondary building material is limited, but the required quality will allow a safe and easy aftercare-free disposal of these materials.

The efforts for metal recovery from filter ashes – seen today in Switzerland – will find an echo in other countries and will pave the way for a tendency to inertize these actually problematic residues and to achieve their easier final disposal.

The today and also in future problematic residues from WTE processes are the scrubbing residues. For the time being there is no convincing solution for the management of these highly water soluble substances other than their disposal in old salt mines – may it be acknowledged as utilization or not. This option is for most countries not accessible. Hence further research is necessary, but from a chemical perspective most of the water soluble species, for example, alkali halogenides, cannot easily be stabilized against leaching. Hence, eventually other practicable approaches have to be considered.

Either recovery of chlorine and extended utilization of sulfur compounds are improved, or these species are discharged into their natural sink – the ocean. Of course, such strategy would need a significant change

in our present gas cleaning “philosophy.” Which technology will be preferred in future is open; however, there is great hope that a sustainable solution will be found.

Bibliography

Primary Literature

- de Fodor E (1911) Elektrizität aus Kehrriecht. K.U.K. Hofbuchhandlung von Julius Benkö, Budapest
- Ude O (1934) Zur Frage der Müllschmelzung. Die Städtereinigung, 26, No. 17 & 18
- International Ash Working Group (IAWG): AJ Chandler, TT Eighmy, J Hartlén, O Hjelmar, DS Kosson, SE Sawell, HA van der Sloot J (1997) Vehlows (1997), Municipal solid waste incinerator residues. Elsevier, Amsterdam/Lausanne/NewYork/OxfordShannon/Tokyo
- Vogg H, Hunsinger H, Merz A, Stieglitz L, Vehlows J (1991) Head-end-Techniken zur Dioxinminderung VDI Berichte 895:193
- European Council (1999) Council directive 1999/31/EC of 26. April 1999 on landfill of waste. Official Journal of the European Communities, 16.7.1999, L182
- European Council (2003) Council decision 2003/33/EC of 19 December 2002 establishing criteria and procedures for the acceptance of waste at landfills pursuant to Article 16 of and Annex II to Directive 1999/31/EC. Official Journal of the European Communities, 16.1.2003, L11/27
- Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (2009) Verordnung zur Vereinfachung des Deponierechts. Bundesgesetzblatt I 2009, 900
- LAGA (1994) Merkblatt Entsorgung von Abfällen aus Verbrennungsanlagen für Siedlungsabfälle, verabschiedet durch die Länderarbeitsgemeinschaft Abfall (LAGA) am 1. März 1994
- CEN (2002) European Committee for Standardisation: CEN/T – Published Standards. <http://www.cen.eu/CEN/Sectors/TechnicalCommitteesWorkshops/CENTechnicalCommittees/Pages/Standards.aspx?param=6273&title=CEN%2FT>. Accessed March 2011
- Schneider J, Vehlows J, Vogg H (1994) Improving the MSWI bottom ash quality by simple in-plant measures. In: Goumans JJM, van der Sloot HA, Aalbers ThG (eds) Environmental aspects of construction with waste materials. Elsevier, Amsterdam/London/NewYork/Tokyo, p 605
- Bergfeldt B, Däuber E, Seifert H, Vehlows J, Dresch H, Mark FE (2000) Rostaschenqualität nach Mitverbrennung der Shredderleichtfraktion in Abfallverbrennungsanlagen. Müll und Abfall 32:138
- Vehlows J, Rittmeyer C, Vogg H, Mark F, Kayen H (1994), Einfluß von Kunststoffen auf die Qualität der Restmüllverbrennung, GVC-Symposium Abfallwirtschaft, Würzburg, 17.-19.10.1994, Preprints, 203
- Köster R, Vehlows J (1998) Organische und anorganische Kontaminanten in Müllverbrennungsschlacken. FZK-Nachrichten 30:139
- Vehlows J, Bergfeldt B, Hunsinger H (2006) PCDD/F and related compounds in solid residues from municipal solid waste incineration – a literature review. Waste Manag Res 24:404
- Johnke B (1995) Schlackeverwertung und -entsorgung unter Beachtung der Vorgaben gesetzlicher und technischer Regelungen, VDI Bildungswerk, Seminar 43-76-03
- Fiedler H (1996) Sources of PCDD/PCDF and impact on the environment. Chemosphere 32:55
- Pfrang-Stotz G (1992), Mineralogische und geochemische Untersuchungen an Müllverbrennungsschlacken, Intern. Kongress für Umwelttechnologie und -forschung im Rahmen der Europäischen Messe für Umwelttechnik, Basel, CH, 5.-7. 10. 1992, Proc. Block 3, 33
- Pfrang-Stotz G (1993) Gutbett-Temperatur-Bestimmungen an Müllverbrennungsschlacken unter besonderer Berücksichtigung mineralogischer Untersuchungsmethoden. Beihefte zum Eur J Mineral 5
- Zwahr H. (2006) Eigenschaften mineralischer Abfälle, Stand der Aufbereitungstechnik und Untersuchungsverfahren von MVA-Schlacken. BMU- Workshop zu den Anforderungen an die ordnungsgemäße und schadlose Verwertung mineralischer Abfälle. 13. und 14. February 2006, Bundesumweltministerium, Bonn. http://www.bmu.de/files/abfallwirtschaft/downloads/application/pdf/abfw_workshop_zwahr.pdf. Accessed March 2011
- Lahl U (1992) Verwertung von MV-Schlacken – durch Optimierung konventioneller Aufbereitung. Müll und Abfall 24:619
- NEN 7345 (1993) Determination of the release of inorganic constituents from construction materials and stabilised waste products. Netherlands standardisation Institute (NNI), Delft
- NEN 7341 (1993) Determination of leaching characteristics of inorganic components from granular (waste) materials. Netherlands standardisation Institute (NNI), Delft
- Sloot HA van der (2005) Horizontal standardization and harmonization of leaching test methods for waste, secondary raw materials, construction materials and (contaminated) soil. <http://www.ecn.nl/docs/library/report/2001/rx01009.pdf>. Accessed March 2011
- Dijkstra JJ, van der Sloot HA, Comans RNJ (2006) The leaching of major and trace elements from MSWI bottom ash as a function of pH and time. Appl Geochem 21:335
- Schweizerischer Bundesrat (1990) Technische Verordnung über Abfälle (TVA) vom 10. Dezember 1990 (Stand am 1. Jan 1993)
- Bergfeldt B, Däuber E, Vehlows J (1998) Norm und Wirklichkeit - zur Aussagekraft standardisierter Elutionstests. Terra Nostra, Schriften der Alfred-Wegener-Stiftung, Nr. 3

27. Pfrang-Stotz G, Reichelt J (1996) Müllverbrennungsschlacken. Mineralogie, Elutionsverhalten und technische Eigenschaften. Baustoff Recycling und Deponietechnik 12:4
28. Sauter J (2000) Vergleichende Bewertung der Umweltverträglichkeit von natürlichen Mineralstoffen, Bauschutt-Recyclingmaterial und industriellen Nebenprodukten. Diplomarbeit am Institut für Straßen- und Eisenbahnwesen der Universität Karlsruhe (TH)
29. Krass K, Mesters K (1993) Verwertung von Müllverbrennungsasche im Straßenbau. VGB Kraftwerkstechnik 73:841
30. Eurostat (2011) Waste statistics. <http://epp.eurostat.ec.europa.eu/portal/page/portal/environment/data/database>. Accessed March 2011
31. CEWEP (2011) Country reports. <http://www.cewep.eu/information/data/subdir/index.html>. Accessed March 2011
32. Vehlows J. (1995) Reststoffbehandlung – Schadstoffsenske "Thermische Abfallbehandlung". In: Die Thermische Abfallverwertung der Zukunft – Mit 100 Jahren Erfahrung ins nächste Jahrhundert. FDBR-Konferenz, Düsseldorf, 28 Sept 1995, Tagungsband, 56
33. Bergfeldt B, Schmidt V, Selinger A, Seifert H, Vehlows J (1997) Investigation of sintering processes in bottom ash to promote the reuse in civil construction (part 2) – long term behavior, WASCON'97, 4.-6.6.97, Houthem St. Gerlach, NL
34. Faulstich M. (1996) Behandlungsverfahren für feste Rückstände aus der Abfallverbrennung. In: Müllhandbuch, ESV Erich Schmidt Verlag, Berlin, A7125. http://www.muellhandbuchdigital.de/doi/MuA_a7125/dokument.html. Accessed Aug 2011
35. Stahlberg R, Feuerriegel U (1995) Das THERMOSELECT-Verfahren zur Energie- und Rohstoffgewinnung -Konzept. Verfahren, Kosten -, VDI Berichte 1192:319
36. Berwein H-J, Erlecke J (1990) Einsatzmöglichkeiten der Verschmelzung als Homogenisierungsstufe in der thermischen Abfallverwertung. In: Thomé-Kozmiensky KJ (ed) Müllverbrennung und Umwelt 4. EF-Verlag, Berlin, p 225
37. Engweiler J (1997) Von Roll RCP: technological breakthrough in waste recycling. In: Barrage A, Edelmann X (eds) R'97 Recovery, recycling, re-integration congress proceedings, vol V. EMPA, St. Gallen, CH, 74
38. Striegel K-H (1997) Die untertägige Verwertung von Reststoffen in der Bundesrepublik Deutschland. In: Müllhandbuch. EVS Erich Schmidt Verlag, Berlin, A8196. http://www.muellhandbuchdigital.de/doi/MuA_a8196/dokument.html. Accessed Aug 2011
39. Reimann DO (1990) Reststoffe aus thermischen Abfallverwertungsanlagen. Beihefte zu Müll und Abfall 29:12
40. Tobler HP (1989) Konzepte zur Reststoffentsorgung in der Schweiz. VDI Berichte 753:9
41. Hagenmaier H, Kraft M, Brunner H, Haag R (1987) Catalytic effects of fly ash from waste incineration facilities on the formation and decomposition of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofuranes. Environ Sci Technol 21:1080
42. Vogg H (1984) Verhalten von (Schwer-)Metallen bei der Verbrennung kommunaler Abfälle. Chemie-Ingenieur-Technik 56:740
43. Vehlows J, Braun H, Horch K, Merz A, Schneider J, Stieglitz L, Vogg H (1990) Semi-technical demonstration of the 3R process. Waste Manag Res 8:461
44. Schlumberger S (2010) Neue Technologien und Möglichkeiten der Behandlung von Rauchgasreinigungsrückständen im Sinne eines nachhaltigen Ressourcenmanagements. In: Schenk K (ed) KVA-Rückstände in der Schweiz – Der Rohstoff mit Mehrwert Bundesamt für Umwelt BAFU, Bern. <http://www.bafu.admin.ch/publikationen/publikation/01538/index.html?lang=de&lang=de>. Accessed March 2011
45. Lundtorp KL, Jensen DL, Sørensen MA, Mogensen EPB, Christensen TH (2002) Treatment of waste incinerator air-pollution-control residues with FeSO₄: concept and product characterization. Waste Manag Res 20:69
46. Bergfeldt B, Jay K, Seifert H, Vehlows J, Christensen TH, Baun DL, Mogensen EPB (2004) Thermal treatment of stabilized air pollution control residues in a waste incinerator pilot plant. Part 1: fate of elements and dioxins. Waste Manag Res 22:49
47. Baun DL, Christensen TH, Bergfeldt B, Vehlows J, Mogensen EPB (2004) Thermal treatment of stabilized air pollution control residues in a waste incinerator pilot plant. Part 2: leaching characteristics of bottom ashes. Waste Manag Res 22:58
48. Karger R (1990) Verfahren zur Rauchgasreinigung bei der Abfallverbrennung. AbfallwirtschaftsJournal 2:365
49. Kürzinger K, Stephan R (1989) Hydrochloric acid and gypsum (sulphuric acid) as utilisable end products obtained from the KRC process for cleaning flue gases from incinerators. In: Thomé-Kozmiensky KJ (ed) Recycling international. EF-Verlag, Berlin, p 1224
50. Volkman Y, Vehlows J, Vogg H (1991) Improvement of flue gas cleaning concepts in MSWI and utilisation of by-products. In: Goumans JJJ, van der Sloot HA, Albers ThG (eds) Waste materials in construction. Elsevier, Amsterdam, p 145
51. Menke D, Baars BA, Fiedler H (1999) Salzsäure aus Müllverbrennungsanlagen: Produkt oder Abfall? Müll und Abfall 31:490

Books and Reviews

- Christensen TH (ed) (2011) Solid waste technology & management. Wiley, Chichester
- Kranert M, Cord-Landwehr K (eds) (2010) Einführung in die Abfallwirtschaft. Vieweg + Teubner Verlag/Springer Fachmedien Wiesbaden GmbH, Wiesbaden
- European Commission (2005) Integrated pollution prevention and control (IPPC) reference document on the best available techniques of waste incineration. Institute for Prospective Technological Studies, European IPPC Bureau, Seville, Spain, July 2005. <http://www.ipcc.ch>

Waste-to-Energy Ash Management in the United States

FLOYD HASSELRIIS

Hasselriis Associates, Forest Hills, NY, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Ash Residues from Waste Combustion Systems
Composition of Municipal Waste
Handling of Ash Residues
Landfill Disposal
Early Research on Ash Management
Regulatory Aspects of Ash Management
Treatment of Ash Residues
Environmental Impact of Use of WTE Ash
Ash Management Around the World
Future Directions
Bibliography

Glossary

Analysis Laboratory testing to determine composition and properties.

Ash residue Noncombustible solid matter remaining after combustion of municipal or other waste.

Beneficial use Application of ash residues to useful purposes as distinguished from disposal in landfills.

Bottom ash Noncombustible residues collected from the bottom of a combustion process such as a stoker.

Contaminants Minor and unwanted chemicals and salts that may be harmful to health and the environment.

Fly ash Fine solid particles derived from combustion and carried by gaseous products of combustion.

Heavy metals Metallic chemical element within the upper range of atomic weights, such as the lead and cadmium contained in ash residues.

Landfilling Placing municipal or other wastes upon the land.

Leaching Process of extraction of water-soluble metals and compounds by water passing through a solid waste.

Testing Laboratory analysis of ash residues to determine composition and properties.

Treatment Processing of ash residues to change characteristics or remove harmful components to reduce or eliminate contamination of the environment.

Definition of the Subject

Municipal solid waste (MSW) contains organics that can be converted to heat and power and should not be wasted by dumping in landfills. However, waste-to-energy (WTE) ash residues contain substances that have the potential to contaminate the environment. Management of ash residues therefore requires procedures that assure placement in landfills wherein leachate is collected, or treated to minimize the toxicity of the residues, possibly for beneficial use.

Regulations for management of ash residues have developed differently in various countries, in accordance with the physical and political situation and also the costs of the alternatives. Procedures used by WTE facilities in the USA require testing of the combined residues from the stoker, boiler, and emission control system to ensure meeting established federal and state regulations. Most of the “combined” ash produced in the USA is disposed in sanitary landfills as Alternative Daily Cover and for other landfill maintenance needs. In contrast, in many European countries bottom ash is used beneficially, for example, for road construction.

Procedures for managing WTE ash in the USA are described below. A following chapter in this Encyclopedia, by Dr. Jurgen Vehlow, discusses the management of WTE ash in Europe and Japan.

Introduction

In the past and even today in many parts of the world, waste materials have been simply dumped, out of sight and out of mind, without regard for their impact on the environment. Rain falling on these deposits leaches toxic metals, and other hazardous materials that can contaminate the soil, air, and waters. Recycling of usable materials is common practice but cannot recover more than one-half of the waste generated by our society, at the most. Landfilling can be reduced to

a minimum by means of combustion with energy recovery, also called waste to energy (WTE).

Municipal solid waste (MSW) is combusted to produce energy throughout the world, especially in developed countries where landfilling is very costly or prohibited. The energy stored in the hydrocarbons contained in MSW is recovered efficiently by means of combustion with heat recovery. The environmental impact of the combustion gases is minimized by the air pollution control (APC) system of the WTE plant. However, the noncombustible fraction remains as a residue, requiring processing, transportation, and disposal or, preferably, beneficial use. The cost of disposing of WTE residues represents a considerable part of the operating cost of the facility. Therefore, management and disposal of ash residues is an unavoidable and critical component of recovering energy from waste.

Ash residues from the combustion of MSW include the bottom ash, siftings, boiler ash, fly ash, and any other residues of the APC system. In particular, the APC residues contain the volatile metals contained in the raw waste in concentrated form and have the potential to contaminate the environment. Regulations controlling the management of ash residues vary from country to country, reflecting different political and physical situations and the cost of alternatives. In the USA, where over 60% of the MSW is still landfilled in sanitary landfills, all ash residues are usually mixed together ("combined ash"), tested for toxicity, and then disposed of in MSW landfills as daily cover of the working area and other maintenance needs, or in dedicated landfills called "monofills." All these landfills are lined so that the leachate can be collected and treated before being discharged. Bottom ash has also been used beneficially in the USA but to a very limited extent.

In northern Europe, where landfills have been phased out and are more costly and even prohibited, WTE bottom ash has been used beneficially, for example, as a substitute for aggregate materials, for use as road base, and for filling trenches carrying pipes for district heating. In Japan, ash residues must be vitrified or otherwise treated so as to permit beneficial use and/or to eliminate the risk of contamination of the environment.

The ash residues from combustion of municipal solid waste represent from 20% to 30% of the weight of incoming waste and 10% of its volume. Since they are generally in a wet condition when disposed, moisture represents 25 or more of their weight. Additives, such as calcium oxide, that are used to capture acid gases and other contaminants in the flue gas add to the weight to be disposed.

The disposal of ash residues must take into account the impact on the environment since soluble toxic metals can be leached out when subjected to rainfall or contact surface or groundwater. Also, reagents used to capture the acid gases may be soluble and influence the solubility of metals. Recognition of these hazards has led to strict environmental controls and requirements for testing of the residues and the leachate. Procedures have been developed to protect the environment and to encourage and permit beneficial uses of the ash, in preference to landfilling.

After passage of the Clean Air Act, the United States Environmental Protection Agency (U.S. EPA) was established. The Resource Conservation and Recovery Act of 1976 (RCRA) empowered the EPA to regulate residues from solid waste incinerators. The U.S. EPA then had to generate research needed to develop and promulgate regulation of waste disposal, particularly management of landfills. Regulations were also developed to assure environmentally acceptable combustion of wastes.

Ash Residues from Waste Combustion Systems

Ash residues are discharged at various locations from the combustion and emission control equipment:

- *Bottom ash* issuing from the discharge of the stoker grate consists of glassy residues, glass, and metallic objects, containing 1–3% carbon. It is usually quenched with water, although it can also be collected in a dry state.
- Stoker grate *siftings* fall through clearances in the grates and are collected with the bottom ash. These may include partially or totally unburned organic matter.
- *Boiler ash*, carried from the grate by the combustion gases, consists of flying particles and condensed metal vapor that may attach to refractory and

water-cooled walls and be caught on boiler surfaces, and has to be removed periodically. It may fall onto the stoker in the bottom ash, or it may be collected in hoppers and be discharged into the *bottom ash*.

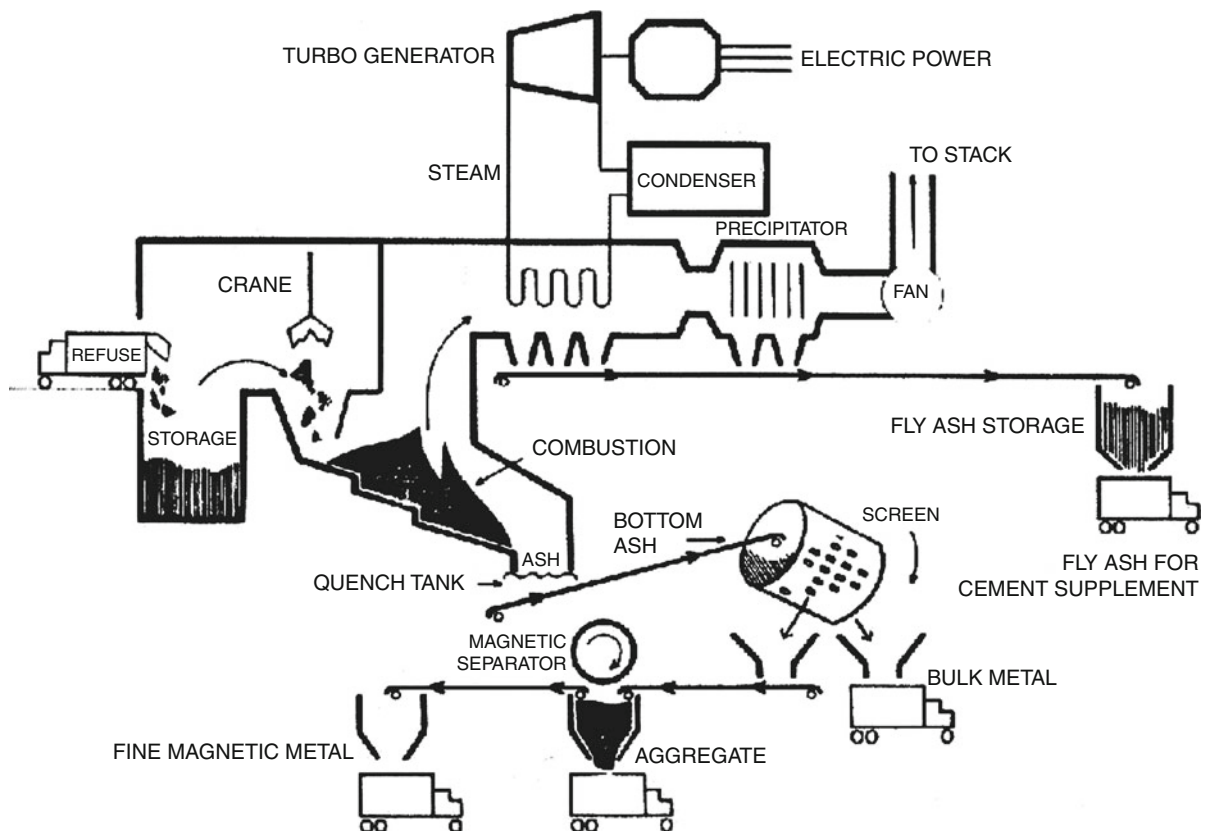
- *Fly ash* is carried by the combustion gases through the furnace, boiler and emission controls. It may be collected in hoppers, by a wet scrubber or by an electrostatic or fabric filter. It may be removed separately or combined with the bottom ash.
- *Scrubber reaction products*, collected at the bottom of a spray-dry or dry lime-injection acid gas scrubber, include fly ash and reacted or partially reacted alkaline reagents such as lime, and activated carbon.
- *Mixed ash* may contain siftings, bottom ash, boiler deposits, fly ash, and scrubber residues.

Figure 1 is a schematic diagram of the Wheelabrator WTE facility built in Saugus, Massachusetts in 1975.

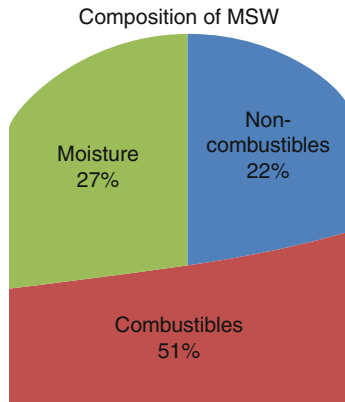
Municipal solid waste (MSW) is fed by crane from the storage pit to the stoker, where it is burned. *Bottom ash* residues fall from the end of the stoker into a quench tank, and are conveyed to a trommel screen which removes bulk metal. The ash drops onto a conveyor, passes under a magnet to remove fine ferrous material, and drops into the aggregate hopper and a transport truck. *Fly ash* from the combustion chamber passes through the boiler, dropping some *boiler ash* into boiler and precipitator hoppers, then through the precipitator which collects the *fly ash* in storage hoppers before being sent out as cement supplement.

Composition of Municipal Waste

Municipal solid waste (MSW) typically contains about 50% combustible matter, 25% moisture, and 25% noncombustible inert materials (Fig. 2). A portion of



Waste-to-Energy Ash Management in the United States. Figure 1
Schematic of Wheelabrator waste management facility, Saugus, MA [33]



Waste-to-Energy Ash Management in the United States. Figure 2

Typical composition of municipal solid waste (MSW) [27]

the metals and glass can be reduced by recycling. The *chemical composition* of MSW is determined by laboratory analysis, after the large inert materials (metals, glass, and ceramics) are removed and accounted for, as performed by ASTM procedures [3]. Table 1 shows a typical ultimate analysis of the combustibles, containing 6.86% inherent ash. The noncombustible major metals constitute 99% of this, mainly aluminum, calcium, sodium potassium, and silicon, as shown in Fig. 3 [27]. Trace metals, totaling about 1% of the noncombustibles are mainly zinc, tin, lead, and chromium.

Table 2 shows the range of quantities of residues generated from combustion of MSW and associated emission controls in 1989. The dry bottom ash and fly ash may constitute 27–39% of the weight of MSW. Residues from various types of acid gas emission controls may add 1–5% [47].

Metals Found in Bottom Ash and Fly Ash

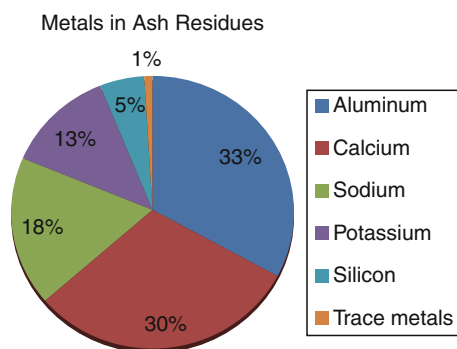
The metals found in combined bottom ash and fly ash, in fly ash and in combined fly ash, and scrubber residues are shown in Table 3 [24]. Lead, cadmium, and zinc exist in high concentrations in the fly ash since they are volatilized at combustion temperatures. Potentially toxic metals in the fly ash, such as lead, cadmium, chromium, and nickel, are mainly derived from pigments, fillers, and inks used in or on paper and plastic products.

Waste-to-Energy Ash Management in the United States. Table 1 Composition of municipal solid waste (MSW) [27]

Ultimate analysis of combustibles	Percent		Metals in ash	Percent
Carbon	35.02			
Oxygen	30.16			
Hydrogen	11.67			
Sulfur	1.88			
Chlorine	0.25			
Moisture	19.16			
Ash	6.86	→	Aluminum	33.21
Total	100.00		Calcium	30.20
			Sodium	17.50
			Potassium	13.12
			Silicon	5.00
			Major metals	99.04
			Zinc	0.400
			Lead	0.310
			Tin	0.150
			Chromium	0.047
			Nickel	0.020
			Cadmium	0.009
			Copper	0.007
			Minor metals	0.960
			(Total)	100.000

Chemical Composition of Ash Residues

The chemical composition of bottom ash and fly ash consists of the minerals found in earth. Fly ash contains greater amounts of the earth metals, the trace metals, lead, cadmium, and zinc, and the highly soluble chloride and sulfate salts. The fly ash/scrubber residues (Table 4) consist mainly of calcium, silica, alumina, and iron oxides. Table 4 compares the concentration of lime (calcium oxide) and other minerals found in air pollution control (APC) residue to that in Portland



Waste-to-Energy Ash Management in the United States. Figure 3

Metals in ash residues from municipal waste [27]

Waste-to-Energy Ash Management in the United States. Table 2 Quantities of combustion and emission control residues

	Quantity of waste, lb/100 lb waste	% of total
Combustion residue		
Bottom ash (slag)	25.0–35.0	90
Filter dust (fly ash)	2.0–4.0	10
Total	27.0–39.0	100
Additional residues		
Wet scrubber residue	0.8–1.5	3–4
Spray-dry scrubber residue	1.6–3.5	6–9
Dry injection residue	2.5–4.5	9–12

Source: Thome-Kozmiensky [47]

cement and flue gas desulfurization (FGD) ash from a coal-fired power plant [28]. Vitriified slag from MSW consists primarily of silica, iron, calcium, sodium, and aluminum [1].

Density of Ash Residues

Combined ash produced by the combustion of MSW has densities ranging from 1,040 to 1,200 kg/m³ at moisture content of 15–25%. In a monofill, combined ash is about 2,200 kg/m³, at about 20% moisture [29].

Waste-to-Energy Ash Management in the United States. Table 3 Metals in combined ash, fly ash, and fly ash/scrubber residues

(Parts per million parts of ash by weight)			
Metal	Combined ash	Fly ash	Fly ash/scrubber residue
Aluminum	17,857	27,500	12,714
Calcium	33,642	64,857	176,428
Sodium	3,828	26,028	1,292
Potassium	3,071	36,928	8,450
Iron	20,428	10,857	3,314
Chloride	928	65,428	164,285
Sulfate	7	33	8
Lead	3,142	22,143	3,257
Cadmium	35	642	160
Zinc	4,107	53,500	9,143
Manganese	534	649	365
Mercury	ND	3	73

Source: Forrester [24]

Waste-to-Energy Ash Management in the United States. Table 4 Comparison of MSW APC waste with Portland cement

Component	MSW/APC ^a residue	Portland cement	Utility ^b FGD ash	MSW ^c slag
CaO	62–67	24	37	14
SiO	18–24	28	24	61
Al ₂ O ₃	4–8	14	6	5
Fe ₂ O ₃	1.5–4.5	5	3	

^aMSW/APC residue: fly ash plus spray-dry acid gas control

^bUtility scrubber residue (From Goodwin [28])

^cSlag obtained by fusing MSW ash residues [1]

Densities of fine and coarse fly ash from a WTE using a dry lime-injection scrubber and baghouse range from 1,900 to 2,600 kg/m³.

Moisture Content

Moisture content of ash residues can range widely, from 15% to 57%, depending on whether a semidry

ash discharger or a water quench tank is used. Excessive moisture content increases disposal costs. Maintaining minimum but sufficient moisture essentially eliminates dust liberation and fugitive dust problems. Controlling moisture can optimize landfill density. Dry fly ash needs to be conditioned with water to eliminate dusting.

Particle (Grain) Size Distribution

Typical size distributions of bottom ash and fly ash, an important factor in disposal and beneficial use of ash residues, are shown in Fig. 4. Typically, these distributions plot as straight lines on log-log paper. The “knee” in the fly ash curve indicates the presence of agglomerated fly ash. The size distribution of bottom ash corresponds well with specifications for aggregate materials [7].

Solubility of Metals in Water

Water containing high concentrations of dissolved metal compounds has the potential to contaminate the environment and drinking water supplies. The solubility of the trace toxic metals depends upon their chemical form and on the acidity or alkalinity (pH) of the water with which they are in contact. The

acidity or alkalinity of ash residues in the presence of water, as measured by pH, generally ranges from neutral (pH = 7) to alkaline (pH = 9). When large amounts of unreacted lime are present, higher alkalinities, up to pH = 11, may be encountered.

Table 4 compares the amount of lime (calcium oxide) found in air pollution control (APC) residues with that found in Portland cement. Hydroxides, sulfates, and chlorides are more soluble and more readily available for leaching than oxides, silicates, and carbonates. The sulfate and chloride forms of lead and cadmium, found mainly in fly ash as the result of reaction with the sulfur and chlorine in MSW, are highly soluble.

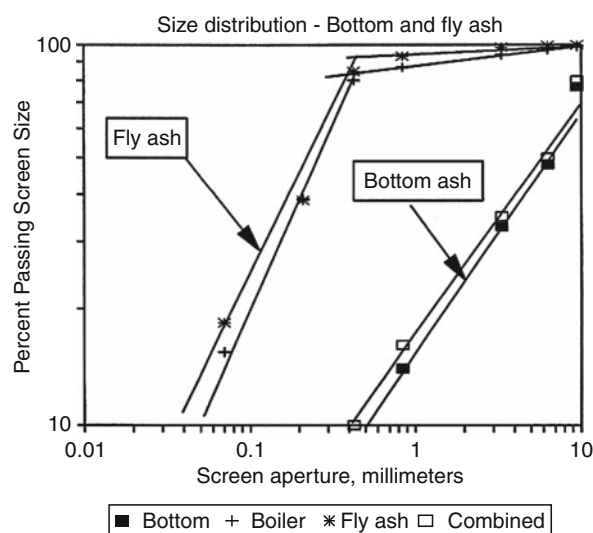
Most metals are only slightly soluble in water at a pH range from 6 to 9. However, under acidic conditions, such as pH of 5 or less, solubility increases rapidly. While most metals are not soluble under alkaline conditions, lead becomes highly soluble at a pH greater than 10. Since lime has a pH of 12, large quantities of unreacted lime in the fly ash from acid gas control devices can increase the concentration of lead in the solution. See Fig. 5.

Soluble Salts in Fly Ash

In general, fly ash contains a large fraction of soluble salts, due to the sulfur and chlorine in the waste. Alkaline reagents such as caustic soda and lime, used as scrubbing agents for acid gas control, react with the acid gases to form soluble salts such as sodium and calcium chloride and various sulfates. Table 5 shows the analysis of the fly ash and scrubber residues from a spray-dry scrubber and the soluble fraction which totals 56% of the fly ash, primarily chlorides and sulfates [40].

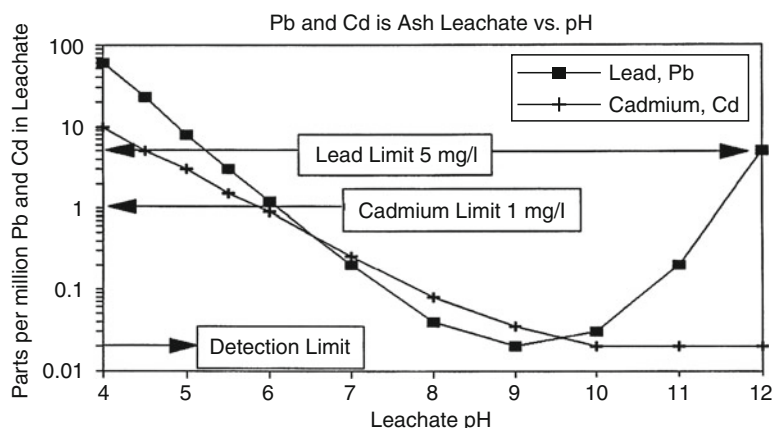
Permeability of Ash Residues

Permeability is the property of a material that measures the velocity at which water will pass through the material, usually reported as centimeters per second (cm/s). The permeability of compacted ash has been measured to be from 1×10^{-6} to 1×10^{-9} cm/s. For comparison, landfill liners are typically required to have a permeability of 1×10^{-7} cm/s. It is possible to prepare ash for use as a landfill liner, thus meeting this specification by proper compaction, and by adding Portland cement and/or lime [24].



Waste-to-Energy Ash Management in the United States. Figure 4

Size distribution of particles in bottom and fly ash [7]



Waste-to-Energy Ash Management in the United States. Figure 5

Solubility of lead and cadmium as a function of pH of the solution. The limits set by the U.S. EPA are indicated: 5 mg/L for lead and 1 mg/L for cadmium. The detection limit is 0.02 mg/L (From Donnelly et al. [12])

Waste-to-Energy Ash Management in the United States. Table 5 Solubility of fly ash residues of dry scrubber [40]

Residue	Weight percent	Soluble percent
Fly ash	40	
CaCl ₂	27	27
CaSO ₃ ·0.5 H ₂ O	20	20
CaSO ₄ ·2 H ₂ O	5	5
CaF ₂	2	2
Lime inerts	2	
CaCO ₃	2	
MnCl ₂	1	1
Heavy metals	$\frac{1-2}{100}$	$\frac{1}{56}$

Handling of Ash Residues

Wet quench systems cool the residues and permit them to be removed from the quench tank by means of *drag conveyors*. Ash residues commonly have a moisture content of about 50%.

Drag conveyors extract ash residues from water-filled quench tanks. They consist of flytes attached to moving chains carried over sprockets, configured to

carry the ash residues up a slope so that they can be conveniently discharged.

Semi-wet systems quench the residues with water, but employ *mechanical dischargers* to push the residues out with a minimum amount of moisture. The heat remaining in the residues serves to drive off moisture so that the discharged residues may have moisture content close to 25%.

Dry removal of the ash residues makes it possible to remove the fine ash component by screening, leaving a useful granular aggregate and a relatively clean ferrous product.

Screw conveyors move powder-like dry ash from fly ash hoppers to other locations where they are dropped into water tanks, conditioning (wetting) devices for damp discharge, or into dry collecting containers.

Vibrating conveyors convey damp ash residues from ash dischargers to processing devices or receiving containers.

Vibrating screens separate ash residues into various size fractions, to remove rejects and produce useful fractions.

Grizzly screens remove oversize objects such as wood, bulky appliances, wheels, and miscellaneous metal objects from the ash residue. Vibrating rails separate out the oversize objects.

Trommels are rotary screens that remove oversize objects and clean ash residues to obtain more uniform products.

Magnetic separators used to recover ferrous metal are usually placed after grizzly screens to reduce interference by oversize objects.

Pneumatic conveyors transfer fly ash from ash hoppers to remotely located containers. Blower air and fabric filters are used to separate the fly ash from the transport air.

Handling and Storage

Handling and disposal of ash residues must not allow contamination of the environment by fugitive dust or by leaching into the environment or water supplies. Dusting is minimized by keeping the residues in a moist condition [30].

Stored ash residues must be properly managed to prevent unacceptable discharge of dust or leachate. Runoff and leachate must be collected, supervised, and properly disposed of so that soluble metal compounds and salts do not contaminate the environment.

Storing of ash allows chemical reactions that bind the metals to take place, reducing leaching potential. Rainwater percolating through the pile cleans the ash while slowly leaching out the available (soluble) metals.

Ash containers must be drainable to recover leachate, and watertight to prevent leachate from running off into the environment.

Transportation of ash residues must be in covered containers to prevent dust from escaping into the environment. Containers, vehicles, and roads must be washed if contaminated with ash. The wastewater must be disposed of properly or recycled back to the ash quench tank.

Processing

Ash residues can be processed at the waste-to-energy facility in order to reduce the rate of release of contaminants into the environment, facilitate disposal, improve the quality of the residues, remove valuable useful or harmful materials, and to prepare portions of the ash for beneficial use. The residues can be treated by washing, chemical treatment, or the use of additives and specific chemicals in order to retain, remove, or immobilize potentially toxic compounds.

Screening processes remove unwanted oversize and undersize components and separate the ash into usable products, including aggregate for use in construction.

Washing processes can provide clean aggregate materials and ferrous metals for beneficial use, and remove cementitious fly ash. Wash water can be processed and recirculated. The blowdown stream must be treated or evaporated to remove/recover and render harmless the dissolved metals and salts [20, 35].

Metal Separation

Metals are recovered not only for their commercial value, but also to minimize problems in beneficial use of the ash residues. As much as 15% of the bottom ash from mass-burn facilities is ferrous material that can be extracted magnetically. Conventional systems recover 60–70% of the ferrous metals and roughly 30% of aluminum, copper, and stainless steel, the nonmagnetic metals.

Tumbling of the ferrous product in a trommel can separate contaminants to improve quality.

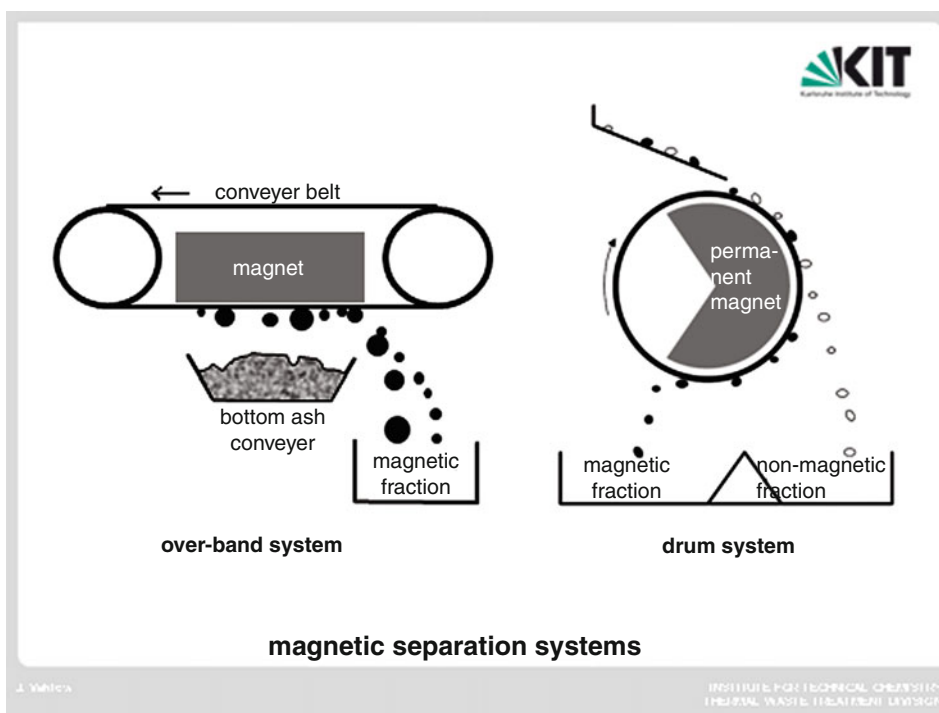
Ferrous metals can be recovered and removed from bottom ash by magnetic separation systems, using either over-band or drum systems (Fig. 6). Non-ferrous metals are recovered by eddy current separators (Fig. 7) [50].

After ferrous separation and screening, bottom ash residues have been used for fill and road base.

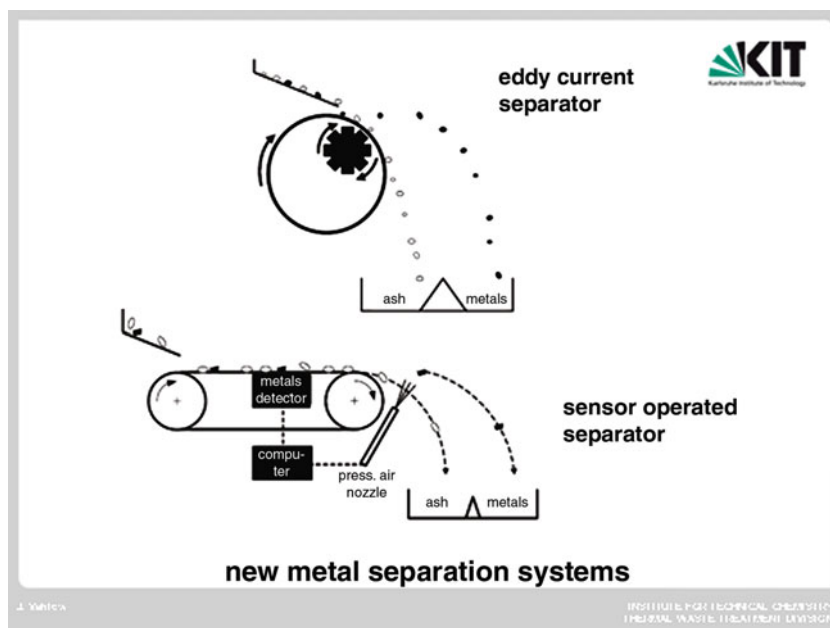
Treatment

Various methods of treatment of the bottom ash may be used to reduce the amount of leachable metal and salt concentrations, and thus render the ash more environmentally acceptable, as well as improve the chemical and physical stability and durability of the product so that it can be used for a variety of purposes. Treatment methods include ferrous separation and compaction, and various methods that modify release rates of the metals by chemical and physical changes, including solidification, stabilization, and encapsulation; addition of Portland cement, phosphates, waste pozzolans, and bituminous materials; washing and chemical treatment, thermal treatment, and vitrification.

Encapsulation in asphalt for use in bituminous paving mixtures and in cement serves to minimize the leaching of metals and salts from the product. Cement blocks and special forms can be made using aggregate processed from ash residues.



Waste-to-Energy Ash Management in the United States. Figure 6
Magnetic separation systems for ferrous metals [50]



Waste-to-Energy Ash Management in the United States. Figure 7
Metal separation systems for nonferrous metals [50]

Lime and/or Portland cement can be mixed with fly ash to encapsulate the toxic metals and/or render them insoluble [38].

Phosphate treatment converts soluble lead compounds to insoluble phosphates, thus immobilizing the lead and reducing leaching to acceptable levels [14].

Carbonic acid absorbed from stack gases can convert soluble lead compounds into relatively insoluble lead carbonates [46, 51]. Heavy metals can be removed from fly ash by using carbonic acid recovered from flue gases, producing insoluble carbonates.

Chemical processing systems can process the ash residues while treating the flue gases and minimizing or eliminating water discharges. Calcium hydroxide specifically removes hydrogen chloride (HCl) and mercury; sodium hydroxide removes sulfur dioxide (SO₂); and ammonia (NH₃) can remove nitrogen oxides (NO_x). Hydrochloric acid is used in a primary gas scrubber to remove most of the cadmium and a major portion of the lead content of combined ash. Washing may especially be justified for treatment of fly

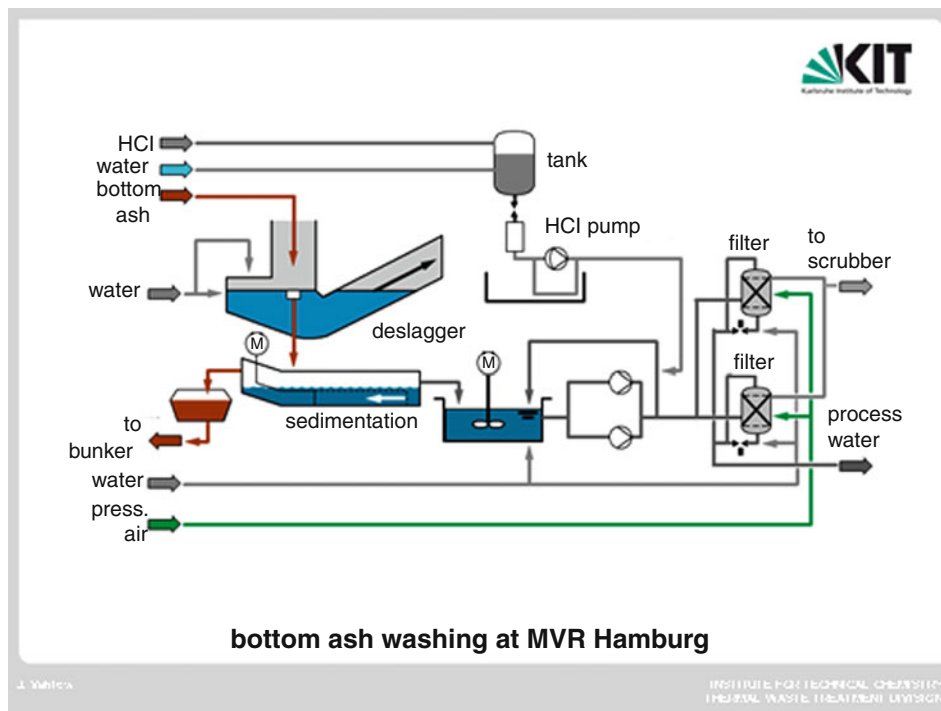
ash, which can contain as much as 50% soluble salts in facilities employing acid gas controls.

Vitrification of bottom ash, mixed ash, or fly ash can be employed to obtain a high-quality glassy frit from which the trace toxic metals such as lead and cadmium do not leach out [10, 11]. Thermal treatment can produce glassy materials, or sintered and ceramic material [51].

Washing of Bottom Ash and Fly Ash

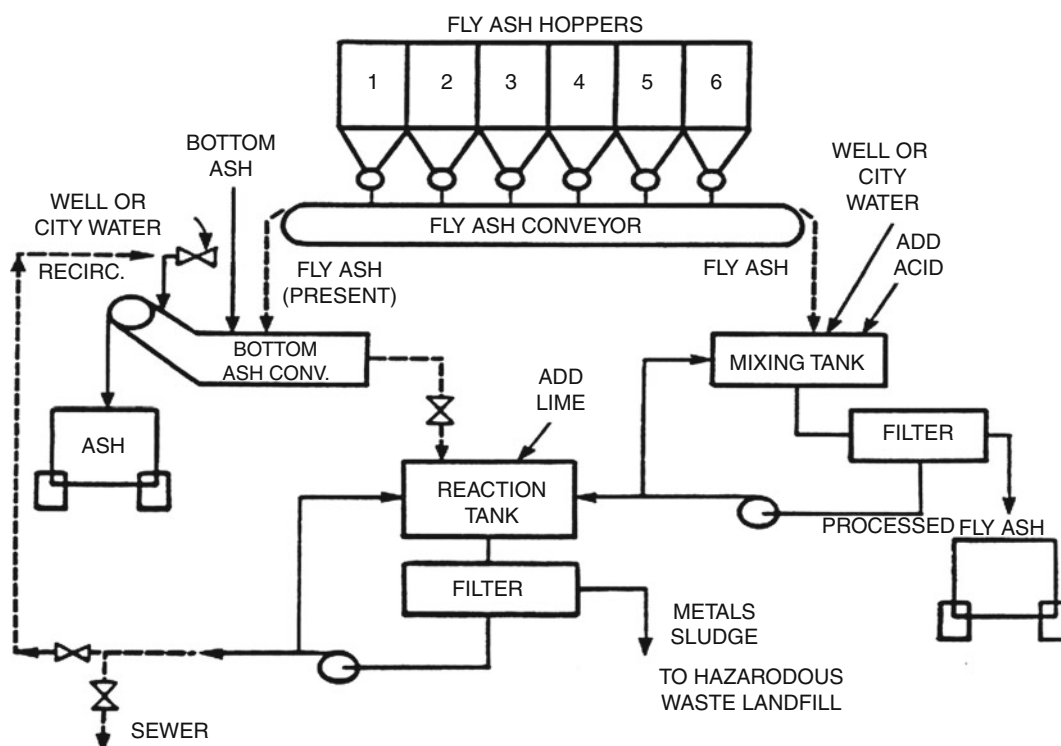
The bottom ash washing system shown in Fig. 8 accepts the bottom ash falling off the grate into a water-filled tank which separates out the slag and drops the residue into a sedimentation tank from which it goes to a bunker. The sediment goes to a treatment tank where HCl is added to force sedimentation of the soluble metals. Recirculated water is supplied to carry off the sedimented metals through the filters which recover the metals.

Figure 9 shows a system which can treat boiler ash and fly ash. The ash from the boiler hoppers is washed



Waste-to-Energy Ash Management in the United States. Figure 8

Bottom ash washing system with treatment of the water at Hamburg [49]



Waste-to-Energy Ash Management in the United States. Figure 9

Treatment system proposed for removing metal sludge and recovering processed fly ash at Saugus [32]

in the boiler ash tank from which it is dragged out to an ash container, for beneficial use. The wash water with this fine ash is passed to a reaction tank where lime is added to control the pH. The fines in the recirculated water are collected by a filter; the metals are discharged as metal sludge and sent to a hazardous waste landfill. A separate treatment system is provided for the precipitator hopper fly ash, where acid is added to precipitate metals. The fly ash can be collected separately, and sent to a mixing tank where acid is added to maintain pH. The fly ash then is filtered and collected for disposal or beneficial use.

Landfill Disposal

Landfilling of untreated ash is the simplest means of disposal. Ash residues are commonly codisposed of with municipal solid waste (MSW). They have been used to cover MSW as daily cover or as final cover. For various reasons it may be preferable to place ash residues in separate or dedicated cells, called ashfills or monofills.

Monofills or Ashfills

Monofills, or ashfills, are used for disposal of ash residues in cells separate from MSW. Placing ash residues in ashfills has the advantage that a solid, relatively impervious mass is created, over which trucks can drive as soon as it is placed. Ash residues have a high density and low permeability, which minimizes the need for leachate collection and treatment [29].

Compacting ash residues can increase the density to as high as $3,400 \text{ kg/m}^3$, as compared with about $1,900 \text{ kg/m}^3$ of uncompacted ash. Another benefit of compaction is the potential for reducing the permeability to as low as 1×10^{-6} to $1 \times 10^{-9} \text{ cm/s}$. Ashfills are so impervious to water that only a small fraction of rain falling on the top surface is able to penetrate the fill; 90% or more will run off, without leaching much of the soluble material in the ash. It is important to provide effective runoff collection systems since the water may be significantly contaminated, especially when the ashfill is in the process of being filled, prior to capping [29].

Codisposal with MSW in Landfills

When ash residues are landfilled together with raw MSW, the acids generated by decomposing MSW could increase concentrations of soluble toxic metals in the collected leachate, requiring more stringent containment, leachate treatment, and groundwater monitoring [25]. On the other hand, the alkalinity of the ash has the ability to neutralize (buffer) the acids, thus reducing acid leaching.

Liners and Containment

Liners are provided in landfills and ashfills to contain the ash residues and to minimize or eliminate leachate penetration into the surroundings, as well as to provide means for collection of leachate and monitoring for indications of leakage. Only a small fraction of the rain falling on an ashfill can percolate through to the leachate collection system. The remainder is runoff that must be collected and properly managed [24].

Ash residues containing mixed bottom ash, fly ash, and acid gas scrubber residues have cementitious and compaction properties that make them relatively impervious to the penetration of leachate, especially if moisture and lime content are optimized. Completed cells can be covered with plastic liners, ash residues, or other relatively impervious materials to essentially eliminate the generation of leachate after the ashfill cell is closed [29].

Ash residues have been used as lining materials for landfills, in lieu of costly clay liners. To prepare the ash for use as a landfill liner, Portland cement can be added at the landfill at 6–10% plus lime at 6–7% by weight. If the fly ash component of the ash residues contains excess lime from the acid gas scrubber, less lime will be needed.

Leachate Disposal and Treatment

Rainfall on the landfill creates runoff and moisture, as it passes through the landfill, and generates leachate which contains various contaminants.

The leachate may be discharged or trucked to wastewater disposal plants if found to be acceptable, or it may require treatment before such disposal. The leachate from fly ash and from mixed bottom ash and fly ash contains substantial amounts (roughly 50%) of

Waste-to-Energy Ash Management in the United States. Table 6 Ranges of leachate concentrations of inorganic constituents from monofills [45]

Constituent	Concentration (CORRE study), mg/L	EP toxicity maximum allowable limit, mg/L	Primary drinking water standard, mg/L
pH	5.2–7.4		
Arsenic	nd–0.400	5.0	0.05
Barium	nd–9.22	100	1.00
Cadmium	nd–0.004	1.0	0.01
Chromium	nd–0.032	5.0	0.05
Copper		1.00	
Iron		0.30	
Lead	nd–0.054	5.0	0.05
Manganese	0.50		
Mercury	nd	0.2	0.002
Selenium	nd–0.340	1.0	0.01
Silver	nd	5.0	0.05
Zinc			5.00
Chloride			250

soluble salts resulting from the removal of acid gases by the emission controls that are required. It has been described as being similar to salt water [36]. The salt content may be more likely to require attention than the low concentrations of soluble metals.

Table 6 shows actual leachate concentrations measured at various ashfill sites. The concentrations of the toxic metals were far below the EPA toxic limits. In most cases, the concentrations of the regulated toxic metals were close to the U.S. EPA drinking water limit [8, 45, 48].

Salinity of Leachate from Ash Residues

The effect of highly salty leachate on the environment has been studied. The salinity of leachate is measured by electrical conductivity. Soils producing leachates that have an electrical conductivity greater than 16 mhos/cm are classified as saline, causing interference

in the uptake of water by plants. Column simulations have shown that after 20 years of leaching, ash residues having an initial leachate conductivity of 21 mhos/cm were reduced by simulated annual acid rainfall to 8 mhos/cm, a level having relatively little impact on most plants [9].

Neutralizing Capacity: Ash/Acid Rain Mass Balance

The soluble toxic metals are only slowly released due to the presence of alkaline materials that provide powerful buffering against MSW-produced acids and the low quantity of weak acids in acid rain. It has been estimated that the effect of acid rain would be resisted for over 1,000 years. Long before this time, the leachable materials would presumably have been removed [31].

Early Research on Ash Management

Wheelabrator built a waste-to-energy plant in Saugus, Massachusetts in 1975, with the concept of total management of the ash residues (Fig. 1). Wheelabrator sponsored research programs at the Agriculture Center of the University of Massachusetts, at Amherst. [21, 32, 34]. In testing use of the ash as a soil conditioner it was found that the ash could enrich the soil, producing faster growth. Tests were run on the leaching of metals from the landfill, which was exclusively combined ash.

A comprehensive testing program was developed to understand the fate of lead and cadmium deposited in the course of flue gas passage through the boiler and the electrostatic precipitator, using samples taken from the dust hoppers of the boiler and the precipitator, and samples from the ash residue as it emerged from the plant [34].

The volatile metals, lead and especially cadmium, were found to collect on the boiler and precipitator surfaces. The suspended boiler tubes were shaken to clean the boiler sections once during each 8-h watch, dropping the metals-containing fly ash into the hoppers. Samples collected hourly revealed a “spike” in the metals content of the fly ash due to tube cleaning, which was reflected in the combined ash residues discharged from the plant. To obtain a true average of the lead content of the residues, a sufficient number of samples must be taken at regular intervals and these samples must be thoroughly mixed.

The leaching behavior of the ash residues was studied. It was discovered that the addition of phosphates to the residues would convert the lead to phosphates that are insoluble, inhibiting the leaching of lead and resulting in passing the Threshold Characteristic Leaching Procedure (TCLP) required by the U.S. EPA [19].

Extensive testing of leaching from landfills proceeded elsewhere: testing MSW landfills, monofills of ash only and codisposal landfills. California regulators concluded that the leaching procedure for ash residues should be performed with deionized water, simulating normal rainwater, not the strongly acidified solution of the “Threshold Characteristic Leaching Procedure” (TCLP) of EPA.

Leachates from ash residues at waste combustion plants throughout the country were tested, including testing of leachate from dedicated ashfills as distinguished from mixed landfills. Tests in Oregon found that the leachate from ash monofills was actually close to neutral, that is, pH = 7.0. Leaching of lead and cadmium is minimized at this pH level.

Since fly ash contains high concentrations of the volatile metals, the practice in Europe has been to keep Threshold it separate from bottom ash and dispose of it as a hazardous waste. Then the bottom ash can be used beneficially. In the USA, however, it has been considered to be more practical to mix the fly ash, and any scrubber residues, with the bottom ash, monitor its leaching properties, and apply additives as needed to pass the EPA leaching test.

Actual Leaching of WTE Ash

Several points are clear from a review of the data of field studies that characterized the leachates from ash monofills, TCLP testing of ash from several waste-to-energy facilities, and leaching of beneficially used products containing ash.

- Laboratory leach tests routinely overestimate the potential for constituents of concern to leach from an ash and ash product when compared to actual leaching from ash monofills and ash use in field applications.
- The combined ash of modern WTE facilities routinely passes the requirements of the TCLP test of EPA.

- Metal concentrations in leachates from ash monofills evaluated over time have routinely met ground water standards and often meet drinking water standards.
- Although leachates from combined ashes in monofills have low concentrations of heavy metals, the total dissolved salts concentration may be several orders of magnitude above drinking water standards.
- Levels of dioxins/furans in ash and ash leachates have been extremely low and considered not to be a concern when evaluating the environmental and health consequences of using ash.

Regulatory Aspects of Ash Management

The management, disposal, and beneficial use of ash residues and their products is subject to federal and state regulations which require sampling and analysis of the leaching characteristics of residues from combustion of municipal and other types of wastes.

Federal Regulations

Federal regulations broadly classify wastes into hazardous and nonhazardous categories. The Resource Conservation and Recovery Act of 1976 (RCRA) empowered the U.S. EPA to regulate residues from solid waste incinerators. In 1992, the U.S. EPA Administrator sent a memorandum to EPA Regional Administrators stating that the Resource Conservation and Recovery Act completely excludes ash from municipal waste combustors from regulation as a hazardous waste under Subtitle C as long as it is not characterized as toxic, since ash can be managed safely in solid waste landfills under Subtitle D, Section 3001(i) of RCRA. Prior to this statement, states developed various regulations, many requiring that MWC ash be disposed of in monofills, for ash only (ashfills), employing single liners, as compared with the double liners required for MSW landfills. Leachate collection and treatment are required in both cases. Landfills for hazardous wastes require much more stringent design and operation due to the greater potential hazards and uncertainty of their leachates.

The U.S. EPA has developed test procedures designed to screen wastes to determine their classification. Various states have developed different regulations as to whether

or not ash must be sampled and analyzed prior to disposal. In any case, it is generally the responsibility of the producer of the ash to determine whether it exhibits the characteristic of toxicity in accordance with U.S. EPA procedures [18].

Ash Residue Extraction Leaching Procedures

The following leaching procedures have been applied to determine the characteristics of ash residues under a wide range of conditions to which they might be exposed, and also to discover which procedures might more closely simulate actual conditions of disposal or beneficial use:

- Extraction Procedure Toxicity (EP Tox) Test
- Acid No. 1. Acetic acid extraction fluid at pH of 4.87 to 5.2
- Threshold Characteristic Leaching Procedure (TCLP)
- Column Tests using simulated acid rain
- TCLP Fluid No. 1 (Acid No. 2). Similar to EP Tox
- TCLP Fluid No. 2 (Acid No. 3). Similar to EP Tox
- California Waste Extraction Tests (WET), Uses citric acid
- Deionized Water (Method SW-924), also known as the Monofill Waste Extraction Procedure (MWEP)
- CO₂ saturated deionized water
- Simulated acid rain (SAR)
- ASTM Shake Extraction Procedure using distilled water leaching

The original Extraction Procedure (EP) toxicity test was replaced by the Threshold Characteristic Leaching Procedure (TCLP). These procedures use an acid leaching medium intended to simulate leaching of ash codisposed in a landfill in a proportion of 15% ash and 85% MSW. In the TCLP test, a unit sample of ash residue is immersed in 20 units of a specified acetic acid solution. The acidity of the solution is maintained at a pH value between 4.87 and 5.2 and stirred continuously for a 24-h period. The extract (leachate) is analyzed, and results are compared with the EPA-established limits. The sample would be characterized as hazardous if any of these limits were exceeded [18, 26]. These limits assume that a 100-time dilution would occur before the leachates could reach drinking water standards, hence they are established at 100 times the drinking water standard.

Actual leachate from MSW landfills as well as codisposal and ash residue landfills (ashfills) does not generally have pH values as low as 5.0. Extensive testing of leachate from these various types of landfills shows that they did not contain significant amounts of the toxic metals lead and cadmium, contrary to the results of laboratory tests employing the TCLP method. In other words, the TCLP test does not simulate actual disposal conditions [45].

A comparison of the effects of various laboratory leaching procedures on the cadmium and lead concentrations is shown in Table 7 [26]. Actual leachate concentrations from tests sponsored by CORRE/EPA are compared with the EPA toxic limits. The carbonic acid test was in closest agreement, while the EP toxicity procedure overestimated leaching by two orders of magnitude [45].

In general, leaching column tests more closely simulate actual leachates from landfills, as do tests *using simulated acid rain or carbon-dioxide-saturated water and/or deionized water*. More aggressive leaching tests such as the California WET test and EP or TCLP test, which represent conditions not likely to occur in the environment, serve as the basis for classifying the

wastes as being *potentially* able to produce toxic levels of metals in the leachate.

Leaching column tests show that the soluble metals and salts are gradually removed from the ash as the leachate absorbs and removes them. The larger the quantity of acid in the water, the faster is the rate of removal. In some cases, two to four quantitative washes will have removed essentially all of the lead and cadmium that was soluble at the final pH level. Test borings of one ash pile showed that after several years of natural acid rainfall the leachable metals remaining in the ash pile had been reduced to the non-detectable level.

State Regulations

State regulations must be at least as stringent as federal regulations, but may be more detailed and suited to specific state environments. They require the collection and analysis of ash residue samples on a periodic basis, and require that these samples, on average, pass the prescribed toxicity tests.

Testing of ash residues is required to obtain confidence that there will be no harmful effects on the

Waste-to-Energy Ash Management in the United States. Table 7 Metal concentrations in extracts of MSW ash residues, in mg/L (parts per million) [16, 45]

Facility	Chicago	Sumner	Hampton	Auburn
Cadmium				
WET test	1.6	0.81	1.52	0.18
EP toxicity	0.71	0.24	0.50	0.02
Acetate	0.19	0.52	0.33	0.03
Carbonic acid	0.016	0.012	0.07	0.005
Water	<0.0005	<0.005	<0.005	<0.005
Lead				
WET test	29.0	35.0	46.0	29.0
EP toxicity	5.8	6.4	10.3	3.15
Acetate	0.5	0.28	1.62	4.20
Carbonic acid	0.025	0.004	0.095	0.012
Water	<0.002	<0.002	<0.002	<0.002

Source: Francis and Maskarinec [26]

environment after the residues are disposed of or used beneficially.

Samples of ash residues must represent the stream of ash residues from which they are taken. Mixing the fly ash properly with bottom ash avoids unrepresentative “hot spots.” Aging samples with normal moisture allows the chemical reactions to take place that would occur under the conditions of disposal, such as converting soluble lead chlorides to insoluble lead carbonates.

Frequency of testing is generally regulated, including extensive testing after start-up of the plant, followed by one or more tests per year in order to assure consistent operation [22]. If some of the analytical results are found to be critically close to acceptable limits, more samples may be taken to obtain confidence in the average values.

Regularly time-spaced samples will represent the true average characteristic of the ash residues for the period of time over which they are produced, landfilled, or otherwise used.

Daily average samples are collected at uniform time intervals over the entire day. Daily samples are well mixed and coned and quartered to reduce the sample size and provide several identical samples. Weekly samples should include the entire week; monthly samples each week, and annual averages all months. As analyses are accumulated over long periods, confidence is established as to the true mean, and fewer analyses are required to assure representation of the residue stream.

Analysis of fly ash samples may show that untreated fly ash contains enough soluble lead and cadmium to exceed the toxic limit according to the EP or TCLP test.

Treatment of Ash Residues

If the ash residues should fail the TCLP test, they would have to be sent to a special and more costly landfill. To avoid these costs, and to assure that all products will pass the test, various chemical treatments have been used.

Processes have been developed to recover useful by-products from waste incinerators, and render the remaining residue nonhazardous, thereby eliminating the long-term environmental liability associated with disposing of these materials.

The fundamental mechanisms of phosphate stabilization of divalent metals in MSW combustion scrubber residues have been studied intensively [16].

The most prominent of these is the WES-PHixTM process, in which the ash residue is sprayed with a phosphate solution while being tumbled in a drum. It has been demonstrated that this treatment reduces the solubility of lead which otherwise increases at the highly alkaline conditions such as 11 or 12 pH which result when excess lime is used for acid gas control. With this treatment, the ash residues may be disposed of in dedicated ashfills, and used for beneficial purposes [42]. Other treatments include control or addition of alkaline agents including Portland cement.

Residues from Air Pollution Control Systems

There are two predominant types of air pollution control (APC) systems: (1) wet scrubbers and (2) dry or semidry lime or other alkaline reagent injection systems. Wet scrubbers generally use sodium-based agents, such as $\text{Na}(\text{OH})_2$, producing the highly soluble NaCl salt. In dry/semidry lime-injection APC systems, the calcium hydroxide ($\text{Ca}[\text{OH}]_2$) reacts with the acid gases to form significant concentrations of calcium chloride (CaCl_2) and calcium sulfate (CaSO_4) that comprise the bulk of the APC residue waste stream. The resulting CaCl_2 and the excess lime present in the APC residues are quickly solubilized upon contact with water. Furthermore, since the stoichiometric ratio of lime addition is greater than 1, the APC residues are highly alkaline, and the potential to solubilize amphoteric metal compounds (such as some Pb compounds) is greatly increased. Consequently, leachates from these residues may contain high concentrations of salts and trace metals, such as Al, Cr, Pb, and Zn.

Environmental Impact of Use of WTE Ash

Risk Assessments

Risk assessments to estimate environmental impact have been carried out for various uses of ash residues: combined ash as landfill daily cover; combined ash as final cover (bottom layer); use of treated ash aggregate (TAA) as a roadway base, as a structural fill, as daily and final landfill covers, as an aggregate substitute in

asphalt concrete paving, and reuse and final disposal of paving material containing TAA.

Boiler AggregateTM from SEMASS, Massachusetts, has been demonstrated for use in producing an asphalt product, as an unregulated fill, and for milling and excavation for reuse. Stockpiling, handling, and transporting operations have also been evaluated, as well as in combined ash as a 30% substitution in bituminous pavement [17].

These assessments have covered specific situations, and generally evaluated the noncarcinogenic and carcinogenic effects on receptors from exposure to As, Ba, Cd. Direct and indirect exposure pathways evaluated included inhalation of fugitive dust on- and off-site, incidental ingestion and dermal contact with ash and ash products, residential exposure to soils potentially contaminated with particulates and/or leachates from ash and ash products, and similar pathways. In the case of landfill final cover, exposure pathways also included incidental ingestion and dermal contact with surface water and sediment while swimming in a nearby harbor, and consumption of fish from the harbor. Exposure from ingestion of food grown in soil potentially contaminated with TAA was evaluated. Also included was ingestion of drinking water containing leachate from Boiler Aggregate stockpiles, a road base, and recycled asphalt product pile.

In tests performed in the USA, on a 600-m section of U.S. Route 3 in Laconia, New Hampshire, with MWC bottom ash as 50% of the required aggregate in the binder course pavement, roadway runoff, surface water, and groundwater were monitored. Monolith leaching tests have indicated that release rates of chemical constituents are low and occur at levels similar to those for natural aggregates [16, 43, 44].

Use of Ash Residues for Construction

Ash residues from combustion of MSW have been used as roadway fill and subbase for parking lots, stabilized road base, bituminous paving mixtures, concrete masonry block, and Portland cement concrete. Concern about leaching of metals into the environment has led to extensive research into the characteristics and environmental impact of these practices. Table 8 describes the physical and chemical tests that may be used to evaluate the waste material for reuse [23].

Use of Ash Residues in Asphaltic Mixtures

Asphalt has been found to encapsulate ash residues effectively, reducing leaching potential to acceptable levels, so that the asphalt can be safely used for road construction [41]. The use of ash residue as an aggregate in bituminous-base course construction has been

Waste-to-Energy Ash Management in the United States. Table 8 Tests recommended for use of granular and asphaltic materials

Chemical tests	Physical tests
Elemental composition mineralogy	Moisture content (ASTM d2216)
Acid neutralizing capacity	Percent rejected (>% in)
Distilled water leach test	Organic content (loss on ignition)
Bioavailability leach test	Ferrous content
Threshold characteristic leaching procedure (TCLP)	Particle size distribution (ASTM C136)
Lysimeter leach test	Absorption and specific gravity (ASTM C127 and C128). Unit weight and voids (ASTM C29)
	Moisture density test (ASTM D1557)
	CBR (ASTM D 1863)
	Sodium sulfate soundness of aggregates (ASTM C-88). Los Angeles abrasion test (ASTM C131)
	Unconfined compressive strength (ASTM D2166). Marshall stability of asphaltic material (ASTM D1559)

Source: Lucido [41]

studied since the 1970s. Test sections have been continuously evaluated over periods from 1 to 5 years. The surface pavement and binder courses using combined bottom ash and fly ash from the Saugus facility in Lynn, Massachusetts, constructed in 1975, used 50% ash, 2% lime, 50% natural aggregate, and 13.5% asphalt, and were assessed in 1991 to be still performing well [15]. A section built in Washington, D.C., contained 68.5% residue, 1.5% hydrated lime, 15% sand, 15% limestone, and 9% asphalt. Surface asphalt road sections at the SEMASS facility in Rochester, Massachusetts, and in New Jersey, and ash use in road base and subbase structures at SEMASS and in New Hampshire were evaluated for soil contamination, runoff, and leachate [5, 52].

Portland Cement Treatment

A patented Portland cement-based ash aggregate, McKayanite, using combined ash or bottom ash, has been tested in Florida as landfill cover and as aggregate in road projects, and was found to meet physical standards for road construction materials, showing no adverse effects on groundwater, soil, or ambient air quality. The State of Tennessee Highway Department developed a standard for acceptable use of ash residues as aggregate in road base construction. Progress toward acceptance of ash residues for beneficial use was expected in other states [52].

At the Commerce WTE facility in Los Angeles, California, fly ash from the spray-dry scrubber/baghouse was mixed with Portland cement, and then blended with bottom ash in a cement mixer truck. The treated ash-concrete was poured into roll-off containers and stored for 24 h before transport to the landfill, where it was crushed for use as a subbase for roads at the landfill [13].

Building Blocks and Other Uses

Use of ash residue to produce aggregate material for use in concrete encapsulates the heavy metals and converts them to chemical forms that are essentially insoluble. Leaching tests of concrete blocks made from bottom ash only and mixed bottom ash/fly ash have shown that the rate of leaching of metals is insignificant in both underwater marine environments and in aboveground applications [52].

Cement blocks made from MSW combustion in Montgomery County, Ohio, have been used to construct buildings on the county landfill, after testing according to structural testing protocols, including ASTM tests for strength, Underwriter Laboratories (UL) tests for fire resistance, and TCLP tests. The blocks exhibited somewhat higher-than-normal shrinkage indices. The blocks made from coarser bottom ash released more easily from the molds than did blocks containing fly ash as a component [52].

Processing Ash for 100% Recovery and Use

A processing plant was built to use ash residues from the Nashville Thermal WTE facility, to achieve 100% recovery of products, that is, zero discharge to the landfill. The process involves sorting of the ash as it is received; bulk reduction; ferrous separation in three stages: ferrous processing; aggregate wet treatment; air classification to remove unburned product; and non-ferrous separation, processing, and sorting. The unburned material was returned to the WTE plant. The product was treated by the WES-PHix process, expecting to sell it for uses in construction. The product was finally conditioned by storage in piles. In spite of successful operation for a long period of time, the facility was ultimately closed due to concern about groundwater contamination from the storage piles which resulted from failure to sell the product. The WTE facility (one of the earliest in the USA) was shut down to allow more profitable use of the downtown site [2].

Vitrification

Vitrification of residues from thermal processes produces a dense, grainless, amorphous, glass-like material which contains no organic material and from which inorganic mineral matter does not leach significantly, and which has many beneficial uses. The American Society of Mechanical Engineers (ASME) and the U.S. Bureau of Mines [10, 11] carried out an extensive investigation of vitrification. Tests were performed on combined furnace bottom ash and fly ash from three mass-burn WTE plants, fly ash from an RDF-fired WTE plant with acid gas scrubber, and combined ash from a multiple-hearth wastewater treatment sludge incinerator. The tests were performed in a sealed

submerged-arc electric furnace having a capacity of 1 t/h. Fumes from the furnace were cooled and the deposits analyzed. A fabric filter was used to collect particulate for analysis, and an afterburner was provided to burn off any possible organic matter. The stack gases were analyzed for trace organics and metals.

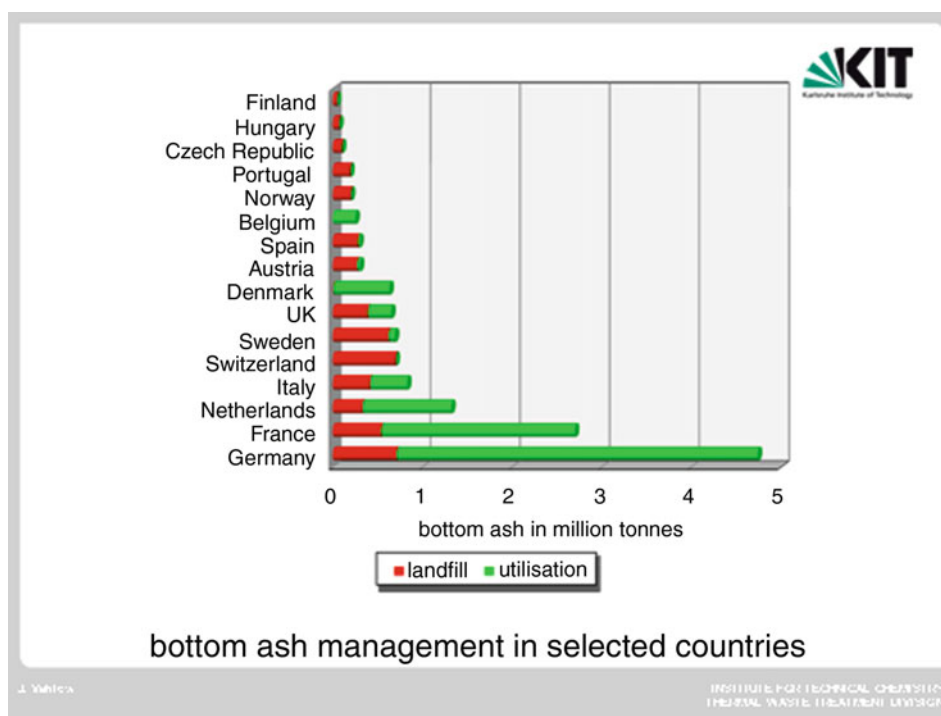
Approximately 83% of the feed weight was converted to vitrified product, and 5% metal was withdrawn, consisting mostly of about 75% Fe, and 8% copper, 6% phosphorous, and 7% silica. The matter, which collects on the top of the charge, contained about 48% copper, 15% iron, and about 27% sulfur. The principal components of the fume solids were NaCl, ZnS, and KCl. They contained 70% of the zinc and 37% of the lead in the furnace feed, and from 7% to 26% silica, varying with the carryover, which depended on the nature of the ash residues. Most of the chlorine in the feed material left the system in gaseous form, since essentially none was found in the solid products.

The vitrified product was tested in accordance with TCLP procedures that showed that the vitrified residues were environmentally benign. The levels of leaching from the samples were in all cases 10–50 times less than the EPA maximum limit.

The cost of vitrification was estimated to range from \$200/t of dry ash residue for a 50 t/day facility to \$115/t for a 300 t/day facility, based on electric power costing \$0.051/kW-h. The components of the cost of vitrification can be roughly estimated to be 5% for drying, 45% for electric power, water, and gas, 30% for labor and maintenance, and about 20% for the cost of capital.

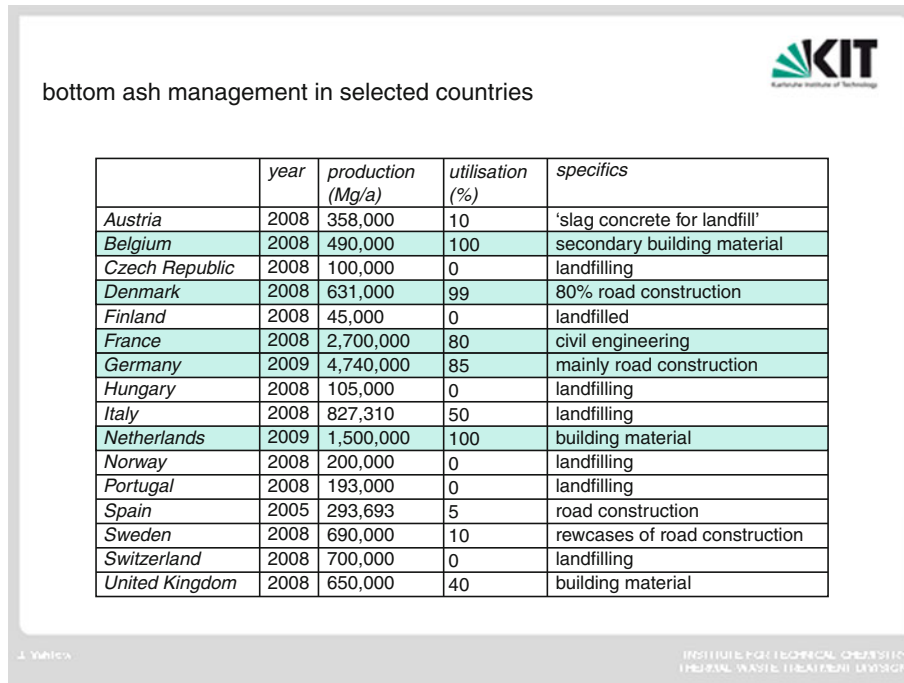
Ash Management Around the World

Worldwide, ash residues represent approximately 20% of the combusted MSW. Therefore, beneficial uses of ash can offer a substantial saving in landfill space and cost [6, 29]. Ash is a resource to be recycled, rather than



Waste-to-Energy Ash Management in the United States. Figure 10

Bottom ash management in selected countries in Europe [49] (KIT – University of the State of Baden-Wuerttemberg and National Laboratory of the Helmholtz Association)



Waste-to-Energy Ash Management in the United States. Figure 11

Bottom ash applications in Europe [49]

a waste to be disposed of into a landfill, provided that utilization is protective of the environment. Many countries have established criteria and procedures for determining acceptable use and disposal options, and continue to support research and development efforts for improved treatment and use technologies [52].

MWC ash is widely used in Europe in road construction as compacted road base; structural fill in wind barriers, sound barriers, and highway ramps; and in asphalt applications [4, 39]. For instance, approximately one-half of the MWC bottom ash generated in Germany is used in road construction. Figure 10 shows the bottom ash utilized in selected countries in Europe versus landfill. Figure 11 shows the quantities and percentages utilized and the specific applications.

Ash management practices in Bermuda, Japan, The Netherlands, Denmark, Germany, France, Sweden, and the UK were reviewed by the U.S. Department of Energy [26]. In Denmark, laws and regulations define how ash residues may be used, and in what quantities, so that environmental impact will be insignificant [37]. In general, in order to produce an environmentally

benign bottom ash material by minimizing the quantity of soluble salts, fly ash is not mixed with bottom ash. The Netherlands, with many years of experience with MWC fly ash, has confirmed its suitability as a substitute for natural aggregate in asphalt road construction. No differences have been found in the leaching of metals between asphalt sections containing MWC fly ash compared to natural aggregates [52].

Future Directions

The many years of research and demonstration have produced a wealth of information which can be used to further sustainability and conservation of materials. The growing number of WTE facilities throughout North America, Europe, Asia, and the world will continue to manage ash residues in accordance with local, state, and federal regulations. The economic and social incentives to make use of these materials and derived products will continue to encourage processing and the marketing essential to favor beneficial use as distinguished from landfilling. Recovery of ferrous metals is

the most rewarding, but next of importance are the nonferrous metals. Ultimately, ash residues should be used for beneficial use in construction as a minimum, but more and more in fabricated products.

Bibliography

Primary Literature

- Abe S (1984) Melting treatment of municipal waste. *Recycling International*, p 173 ff
- Arcani GB (2000) Ash recycling in Nashville, Tennessee. Pub. No. GR-WTE-0108, SWANA, 8th annual North American waste-to-energy conference, Nashville
- ASTM (1989) Standard guide for general planning of waste sampling (D4687). American Society for Testing and Materials, West Conshohocken
- Chandler AJ et al (1997) Municipal solid waste incinerator residue. *Studies in environmental science* 67, The International Ash Working Group (IAWG). Elsevier, Amsterdam, The Netherlands
- Chesner WA (1989) Aggregate-related characteristics of MSW combustion residue, Resource recovery report. In: *Proceedings of the 2nd international conference on MSW combustor ash utilization*, Washington, DC
- Chesner WA (1993) Working toward beneficial use of waste combustor ash. *Solid Waste Power* 7(5):12
- Chesner WH, Collins RJ, Fung T (1988) Assessment of the potential suitability of Southwest Brooklyn incinerator residue in asphaltic concrete mixes. *NYSDA report 90-15*, State of New York, Albany, NY
- Clark RA (1992) Summary of the field leachate generated from Northern States Power Company's Red Wing RDF ash disposal facility, Paper no. 92-47.02. In: *85th annual meeting of the air and waste management association*, Kansas City
- Cundari KL, Lauria JM (1986) The laboratory evaluation of expected leachate quality from a resource recovery ashfill. In: *1986 triangle conference for environmental technology*, Chapel Hill, NC
- DeCesare R, Plumley A (1992) Results from the ASME/US bureau of mines investigation – vitrification of residue from municipal waste combustion. In: *1991 ASME national waste processing conference*, Detroit, MI
- DeCesare R, Plumley A (1992) Results from the ASME/US Bureau of Mines Investigation – vitrification of residue from municipal waste combustion, Paper 92-47.03. In: *85th annual meeting of the air and waste management association*, Kansas City
- Donnelly JE, Jons E, Mahoney P (1987) By-product disposal from MSW flue gas cleaning systems, Paper #87-94A.3., 1987 APCA annual meeting, New York, NY
- Eaton M (1992) The Commerce refuse-to-energy facility ash handling and treatment system. *Chesner Engineering*, 5th international conference on MSW combustor ash utilization
- Eighmy TT et al (1989) Theoretical and applied methods of lead and cadmium stabilization in combined ash and scrubber residues, Resource Recovery Report. In: *Proceedings of the 2nd international conference on MSW combustor ash utilization*, Washington, DC
- Eighmy T et al (1991/1993) The Laconia, NH, bottom ash paving project. Environmental Research Group, University of New Hampshire, Durham
- Eighmy T et al (1997) Fundamental mechanisms of phosphate stabilization of divalent metals in MSW combustion scrubber residues. Pub. No. GRANITE 0105, SWANA. In: *5th annual North American waste-to-energy conference*, Research Triangle Park, NC
- Energy Answers, Energyanswers@Energyanswers.com
- EPA US (1980) Hazardous waste management system: identification and listing of hazardous waste, toxicity test extraction procedure (EP). *Fed Reg* 45(98):33063–33285
- EPA US (1995) Guidance for the sampling and analysis of municipal waste combustion ash for the toxicity characteristic. U.S Environmental Protection Agency, Washington, DC
- Exner R et al (1989) Thermal effluent treatment for flue gas treatment systems in refuse incineration plants from the point of view of residue minimization by recycling. *Recycling International*, pp 1372–1392
- Feder WA, Mika JS (1982) Summary update of research projects with incinerator bottom ash residue. Executive Office of Environmental Affairs, Commonwealth of Massachusetts, Boston
- Fiesinger T (1989) Sampling of incinerator ash. Resource recovery report, p 51. In: *Proceedings of the 2nd international conference on MSW combustor ash utilization*, Washington, DC
- Fiesinger T (1992) Incinerator ash management: knowledge and information gaps to 1987. *Recovery Report 92-6*, NYSDA, Albany, NY
- Forrester K (1989) State-of-the-art in refuse-to-energy facility ash residue characteristics, handling, reuse, landfill design and management – a summary, Resource Recovery Report, pp 167 ff. In: *Proceedings of the 2nd international conference on MSW combustion ash utilization*, Washington, DC
- Francis CW (1984) Leaching characteristics of resource recovery ash in municipal waste landfills. DOE ERD-83-289, No. 2456, Oak Ridge National Laboratory, Oak Ridge
- Francis CW, Maskarinec MP (1987) Leaching of metals from alkaline wastes by municipal waste leachate. ORNL/T-10050, Pub. No. 2846, Oak Ridge National Laboratory, Oak Ridge
- GBB (1990) Studies of waste composition. Gershman, Brickner and Bratton, Inc., Falls Church
- Goodwin RW (1982) Chemical treatment of utility and industrial waste. ASCE National Conference on Environmental Engineering, Minneapolis
- Goodwin RW, Forrester KE (1990) MSW ash field study: achieving optimal disposal characteristics, Paper no. 25094. In: *American Society of Civil Engineers, J Environ Eng Div* 116(5):880–889

30. Hahn J et al (1990) Fugitive particulate emissions associated with MSW ash handling – results of full scale field program, 83 rd annual A&WMA, Pittsburg, PA
31. Harden J, Elander P (1986) Residues from waste incineration – chemical and physical properties, SGI Varia 172. Swedish Geotechnical Institute, Linköping
32. Hasselriis F (1992) Alternative methods of disposal of residues from municipal incinerators. Wheelabrator Technologies.com.
33. Hasselriis F (1993) A comprehensive study of ash residue characteristics and statistical analysis of data. Air and Waste Management Association, Paper No. P-30
34. Hasselriis F (1994) Analysis of data obtained from an historic ash residue leaching investigation. In: ASME national waste processing conference, New York, NY
35. Hasselriis F, Drum D et al (1991) The removal of metals by washing of incinerator ash. In: 84th annual meeting of the air and waste management association, Vancouver, BC
36. Hjelmar O (1982) Leachate from incinerator ash disposal sites. In: International workshop on municipal waste incineration, Montreal, Canada
37. Hjelmar O (1990) Regulatory and environmental aspects of MSW ash utilization in Denmark, Resource Recovery Report. In: 3 rd international conference on ash utilization and stabilization (ASH III), Washington, DC
38. Holland P et al (1989) Evaluation of leachate properties and assessment of heavy metal immobilization from cement and lime amended incinerator residues, vol 1. In: Proceedings of the international conference on municipal waste combustion, Hollywood, FL
39. International Ash Working Group (IAWG) (1997) Municipal solid waste incinerator residues. Elsevier, Amsterdam
40. Lebedur H et al (1989) Emission reduction in the second domestic waste incinerator plant in the RZR Herten. Recycling International, p 1220
41. Lucido SP (2000) The use of municipal waste combustor ash as a partial replacement of aggregate in bituminous paving material. Pub. No GR-WTE-0108, SWANA, 8th annual North American waste-to-energy conference, Nashville
42. Lyons MR (1996) The WES-PHix ash stabilization process, pp 165ff, 17th Biennial ASME SWPD
43. Magee BH, Hahn J et al (1999) Environmental testing of municipal solid waste ash-amended asphalt, pp 103 ff. In: North American waste-to-energy conference, 1999
44. Musselman CN et al (1994) The New Hampshire bottom ash paving demonstration, Laconia, New Hampshire. In: ASME national waste processing conference, New York, NY
45. Roffman HK (1991) Major findings of the US. EPA/CORRE MWC-Ash Study. A&WMA VIP-19, Municipal Waste Combustion
46. Shinn CE (1987) Toxicity characteristic leaching procedure (TCLP), extraction procedure toxicity (EPTox) and deionized water leaching characteristics of lead from municipal waste incinerator ash. Department of Environmental Quality, Portland
47. Thorne-Kozmiensky K (1989) Measures to reduce incinerator emissions. Recycling International, p 1009
48. U.S. EPA (1988) National secondary drinking water regulations, 40-CFR 143. U.S. Environmental Protection Agency, Washington, DC
49. Vehlow J (2002) Bottom ash and APC residue management in Europe. In: Expert meeting on power production from waste and biomass, Espoo, Finland, 8–10 April 2002
50. Vehlow J (2010) Trends in utilizing WTE residues in Europe. WTER 2010 bi-annual meeting, Columbia University, New York
51. Wakamura Y, Nakazato K (1992) Technical approach for flyash stabilization in Japan, Paper no. 92-20.08. In: 85th annual meeting of the air and waste management association, Kansas City, MO
52. Wiles C (1999) Beneficial use and recycling of municipal waste combustion residues – a comprehensive resource document. NREL/BK-570-25841, NREL, Boulder, CO

Books and Reviews

- Hasselriis F (1984) Processing refuse-derived fuel. Ann Arbor Science, Boston (Available from author). ISBN 0-250-40315-3
- Hasselriis F (2002) Waste-to-energy combustion, part 12B ash management and disposal. In: Tchobanoglous G, Kreith F (eds) Handbook of solid waste management, 2nd edn. McGraw-Hill, ISBN 0-07-135623-1
- U.S. EPA (1993) Test methods for evaluating solid waste physical/chemical methods, 3rd edn. SW-846, sampling and testing methods for solid waste. U.S. Environmental Protection Agency, Washington, DC

Waste-to-Energy Facilities as Power Plants

M. A. J. (MARCEL) VAN BERLO
Free University of Amsterdam and AEB-Amsterdam,
Amsterdam, The Netherlands

Article Outline

Glossary
Definition of the Subject
Introduction
Waste to Energy History
Improving Energy Efficiency from Waste
Assessing the Energy from Waste
Future Directions
Bibliography

Glossary

CBA Cost benefit analysis.

CHP Combined heat and power: production of electricity combined with delivery of heat, e.g., for district heating.

EfW Energy from waste, synonym of WtE.

LCA Life cycle analysis.

MSW Municipal solid waste: residential, commercial, and other urban waste. It is the residual part, which is the non-source separated stream.

R1/D10 EU legislative distinction between Reuse and Disposal of waste.

RDF Refuse derived fuel: pretreated waste, mainly by shredding and recovery of some recyclable materials and removal of inert and wet fractions. It consists mainly of packaging paper and plastics.

WTE Waste to energy: Thermal treatment of wastes with energy recovery.

Definition of the Subject

This entry describes the history of WTE and technological developments with regard to increasing the energy efficiency of this process, and ways to assess energy efficiency.

In the history, four distinct stages are described involving improvements on the combustion system, flue gas cleaning, and energy efficiency.

An overview of options to improve energy efficiency is given, with the technical limitations to raise steam parameters, resulting efficiencies, and ways to evaluate the energy performance of WTE installations.

Introduction

Over the last century, Waste to Energy was developed in a few distinct stages. Originally, the objective was to manage urban wastes in order to avoid the spread of diseases. In the second stage, reduction of smoke issuing from the stack became the point of attention. Later, discovery of the effects of invisible pollutants, such as volatile metals and dioxins, in the stack gases led to important advancements of the air pollution control systems of WTE plants. Currently, sustainability and energy and materials conservation have shifted the focus of WTE research and development on obtaining higher energy efficiency and recovery of metals. This

entry describes briefly the history of WTE and presents technological developments with regard to increasing energy efficiency.

Waste to Energy History

Pre–Nineteenth Century

Managing of waste became an urban issue when cities reached a certain size. Nuisance factors, both visual and unpleasant odor effects, led to the adoption of some rules for dealing with waste. In the nineteenth century, landfills were created to deal with the mounting amount of waste and collection was organized either by public or by private companies. The cholera epidemic that started in 1892 in Hamburg, swept over Europe, and even reached America, was a general political trigger for public responsibility for sanitation.

1900–1960: Hygiene

As a response to this political pressure, waste collection, sewer systems, and drinking water distribution were quickly developed in many European cities. Landfills were moved out of the cities and waste incinerators were built in many places. The focus of these incinerations was to dispose of the waste hygienically and reduce the amount of materials that had to be transported to landfills. In order to increase landfill space, in many cases the landfilled waste was often burned in the open. When the waste was burned in dedicated incineration facilities, the focus was on hygiene, and there were good examples of integrated systems of waste collection, transportation, and transfer into the bunker of the incineration plant. These systems were designed to minimize manual contact with the waste and keep it in closed containers to avoid rats, mice, birds, and flies.

However, in this focus on sanitation, there was not much attention paid to atmospheric pollution. A stack plume of black smoke was common and was considered to be a sign of economic activity and prosperity rather than a problem. Also, the bottom ash, and fly ash collected in a simple cyclone, of incinerators was landfilled or spread on roads to harden the surface. The incinerators were generally run as batch operations.

With regard to energy, most of these installations did not provide for energy recovery, but some were heating water for district heating. The very simple combustion systems were inadequately controlled, resulting in large amounts of air used, incomplete combustion causing massive amounts of black smoke and carbon monoxide, and very wide temperature fluctuations. This also led to corrosion problems so that only very simple and small heat exchangers were used to extract some of the combustion heat. As a result, the heat recovery was very small and most of the combustion heat was lost in the flue gasses leaving the stack.

Some incinerators in the first part of the twentieth century were equipped with a boiler for steam production that was used for either industrial use or electricity production. Electrical efficiency was very low, at about 5% in the best cases.

1960–1990: Dust Removal

Over time, the black deposits of the smoke and fly ash of waste incinerators became noticeable and gradually unacceptable. Therefore, standards emerged limiting smoke and dust for the next generation of incinerators. Technology improved in boiler designs, control of the fire, and fly ash separation by means of electrostatic precipitators that resulted in stacks without black smoke. This also enabled the recovery of more heat from the combustion gases and continuous operation of the incinerator and boiler for a period of a month or more. During this period, some cities were provided with a district heating network using hot water. This was done with a simple and inexpensive hot water boiler but did not take advantage of the potential work that can be done with the energy of cogenerating electricity and heat. For Combined Heat and Power (CHP), a steam boiler is needed plus a turbine for electricity production. In this period, CHP installations were mainly small backpressure turbines that were used to produce the heat and electricity.

Steam pressure from the boiler was initially in the range of 10–20 bar, but as designs improved, it was gradually raised for newer installations, reaching a common acceptance of about 40 bar as state of the art at the end of the 1980s. This 40-bar boiler pressure, and 400°C superheater temperature, resulted in an acceptable compromise between excessive corrosion

of the boiler wall and superheater tubes and thermal efficiency achieved.

Also, energy loss in the flue gases was lowered by decreasing the boiler outlet temperature to 250–300°C. The resulting energy efficiencies were gradually improved to about 80% for hot water boilers and from 10% to 18% for electricity producing plants

In this period, some plants in Germany were built with higher steam parameters (e.g., Mannheim: 120 bar/525°C; Düsseldorf: 90 bar/495°C; Frankfurt 60 bar/500°C; Stuttgart 74 bar/525°C). However, these were based on coal-fired boiler concepts. In the case of burning Municipal Solid Waste, the high chlorine-to-sulfur ratio results in high corrosion rates and, also, the fly ash tends to be very stick at the higher temperatures of flue gas and boiler wall. The resulting large downtime periods for repairs and plant availabilities in the range of 70–80% lead to adaptations in the design of boilers that use municipal solid waste (MSW) as their fuel. But changing the steam parameters of a WTE plant is generally not possible without major rebuilding of boiler parts and turbine. Therefore, existing plants could not be retrofitted economically and had to continue operating at low plant availability.

The general consensus developed that 40 bar/400°C was the optimal compromise for the steam parameters, balancing corrosion and associated shutdowns and maintenance with the returns for the electricity produced. The overall efficiency with these parameters was commonly up to about 14–18%, but most operators paid little attention to improving electrical efficiency as the revenues from power production and sale were generally less than a quarter of total revenues.

1990–2005: Chemical Cleaning

At the end of the 1970s, it was discovered that although there was no visible smoke, dioxins and some volatile metals were present in alarming quantities in the flue gas of WTE plants. This raised great concerns and, within a period of about 10 years, political measures were taken to drastically limit the allowable quantities of emissions of dioxins and other critical pollutants such as mercury, other heavy metals, chlorine, sulfur dioxide, carbon monoxide, and particulate matter. The concerns were such that the new limits for dioxins were set at the detection limit, i.e., at about a factor 1,000

lower than average dioxin emissions of the time. In some EU countries standards emerged which were EU-wide implemented following the year 1991. These have by now been taken over by nearly all other countries with WtE installations.

The new standards have led to closure of many of the 1980s WtE plants and the addition of state-of-the-art air pollution control systems in all operating plants. As a result, in a period of ten to fifteen years, the WtE industry, worldwide, has been transformed from a major emitter of pollutants to being one of the cleanest power generators, having stricter limits than almost any other industry. For example, for dioxins, the emissions have been reduced by a factor 1,000 till a maximum limit of 0.1 ng/normalized cubic meter. As a result, by 2005 the issue of emissions from WtE was resolved. Although there remain some local debates at different countries, these are more a remnant of the past image rather than because of actual impacts of WtE plants.

With the focus on cleaning the WtE stack gases, there has been little attention to improvements with regard to energy efficiency. In many cases, there was an actual lowering of energy production, because of the increased use of energy within the plant for gas cleaning.

Some of the existing WtE plants have experimented with adding energy-saving improvements, such as additional heat exchangers in the flue gas (e.g., a second economizer or a condensing heat exchanger) so as to improve the recovery of heat for district heating or to replace internal heat consumption and to increase the plant's electrical efficiency.

Plants that were designed and built in this period generally followed the consensus of the 40 bar/400°C steam parameters. However, on the average, energy efficiency increased somewhat during this period, by reducing boiler outlet temperature (from about 250°C to 180°C), reducing the use of excess air from 1.8 to 1.6 (lowering oxygen content in the stack gas from 12% to about 8%), and simplifying the gas cleaning designs. Few plants that deliver only heat were built, as CHP was commonly employed to produce electricity and heat. When the plant is small compared to the heat demand, backpressure turbines were used. For bigger WtE plants, a vacuum turbine with a steam bleed was used. This allows maximum electricity production while the heat extraction can be controlled according to demand.

Some exceptional WtE plants were built in this period with high thermal efficiencies, such as:

- Falls (Bucks county, PA, USA, 1994): 90 bar/500°C and a boiler outlet temperature of about 180°C resulted in about 28% net electrical efficiency.
- Moerdijk (The Netherlands, 1996): 100 bar/400°C is directly coupled with a Combined Cycle Gas-fired power plant that further superheats the steam and includes it in its cycle. The total efficiency is in the order of 45%, but as natural gas can normally yield 58% efficiency, the net efficiency corresponding to the chemical energy from the waste is about 30%.
- Brescia (Italy, 1998): 61 bar/450°C two lines (plus a third line in 2004 operating at : 73 bar/480°C) and a boiler outlet temperature of 135°C results in about 25.8% net electrical efficiency plus 58% of heat for district heating.
- The Bilbao (Spain), 265,000 t/year WtE plant combusting presorted waste was started in 2004. The WtE steam parameters are 100 bar/330°C, but this steam is then superheated to 540°C in a Combined Cycle Gas Turbine and reheated half-way the turbine. The efficiency is similar to the Moerdijk WtE mentioned above.

Alternative thermal treatment technologies for treating solid wastes (e.g., fluidized bed, pyrolysis, and gasification) have not proven in practice to have as high energy efficiencies as the classical WtE installations that use the common grate technology. This is mainly due to the difficult physical and chemical properties of Municipal Solid Waste (MSW) or Refuse Derived Fuel (RDF) and associated operational problems, as compared to coal or biomass (e.g., wood chips).

2005 to Present: Sustainability

Since the mid-1990s, nickel–chromium alloys have been used to repair and protect corrosion-affected parts in WtE boilers. Especially, the alloy Inconel 625 has proven to be extremely successful in surviving the corrosive conditions encountered in MSW boilers. The gain in plant availability has paid for the high additional investment costs and by the year 2000, nearly all WtE plants were testing or using this alloy. Its high cost however prohibited its use by contractors for new installations. However, by 2005, this alloy was proven to be so effective

that in some new construction projects its use was prescribed for cladding the 1st-pass of the boiler already at construction of the plant. Steam parameters were still mostly around the conventional 40 bar/400°C.

At the same time, sustainability and the greenhouse gas effect came very much in focus, partly due to Al Gore's film, the "Inconvenient Truth." In two legal proceedings in the EU about cross-border transport of waste, it was ruled that the efficiency of the installation receiving the waste was the key criterion for deciding whether this should be allowed. This resulted in an efficiency formula in the EU "Best Available Technology" descriptions. In the new EU Waste Framework Directive, this resulted to the legal distinction between reuse/recovery (R1) and disposal (D10) status. This also meant that energy efficiency determines whether waste can be traded cross border and whether it may be used in the statistics for recovery.

The combination of more attention for energy efficiency and the availability of Inconel 625 as a good performing corrosion resistant material led to the use of higher steam parameters in new plant. Some examples are:

- In Amsterdam (The Netherlands), alongside the existing plant (1993, 800,000 t/year, 42 bar/430°C), the 530,000 t/year Waste Fired Power Plant was completed in 2007. The superior properties of Inconel 625 allowed in 2001 the choice for increasing the boiler steam parameters to 125 bar/440°C [1]. This high pressure with moderate superheater temperature was made possible by a steam-to-steam reheater, using drum steam to reheat the steam halfway the turbine. Energy efficiency was increased further by reducing boiler outlet temperature, reducing the combustion excess air ratio till 1.4 (equivalent to lowering oxygen content in the stack gas from 8% to about 6%), heat recovery from the gas cleaning system, reducing of internal heat and electricity consumption, and improved design of turbine and condenser. The net electrical efficiency of this unit is 30.4% and 34.5% gross.
- In Rüdersdorf near Berlin (Germany, 2009), a 250,000 t/year RDF plant was started with steam parameters of 90 bar/420°C. Also, reheating halfway the turbine is used, albeit with the reheating accomplished by a heat exchanger tube bundle in

the flue gas. The net electrical efficiency of this plant is 29.1% and 31.8% gross.

- In Hameln (Germany), a 200,000 t/year plant was started in 2010 with steam parameters of 41 bar/400°C and external superheating of the steam to 520°C. The excess air ratio is only 1.39. The net electrical efficiency from the energy in the waste is 26.6% and 29.7% gross.
- Biomass combusting plants with reheaters were erected in Bischofferode (Thüringen, Germany, 128 bar/532°C), Leipzig (Germany, 20 MW fluidized bed boiler, fresh wood from forestry, 128 bar/532°C), Eberswalde (Germany, 80 bar/480°C), and Simmering (Vienna, Austria, 120 bar/520°C, 23.4 MWel, gross efficiency of 36.4%). The temperatures and hence efficiencies in these plants can be higher because the biomass combusted contains less chlorine than MSW, which results in less aggressive conditions in the boiler.

The above developments indicate that the focus on sustainable waste management along with favorable legislation and the availability of new corrosion-resistant materials have shifted the frontier of what WTE efficiencies can be commercially achieved. Especially, the general acceptance of using the high-cost Inconel 625 as a cladding on the boiler membrane walls has been a breakthrough in maintaining high availabilities with increasing steam parameters.

Historical Overview

Table 1 shows schematically the different generations of WTE power plants.

Improving Energy Efficiency from Waste

Waste to Energy plants all use a boiler to produce steam and a steam turbine to convert this steam to electricity in a Rankine cycle. The efficiency of this Rankine cycle depends directly on the ratio between maximum and minimum temperature, i.e., the superheated steam and the steam condenser at turbine outlet. The strongly corrosive properties of the combustion gases are limiting the maximum steam temperature that can practically be handled. This explains why WTE power plants cannot achieve the high energy efficiencies of a clean fuel like natural gas.

Waste-to-Energy Facilities as Power Plants. Table 1 Changing paradigms for WtE-plants

Generation	Period	Paradigm	Boiler	Heat efficiency	Electrical efficiency
0	Pre-1900	Landfill	–	–	–
1st	1900–1960	Sanitation , Incineration	Hot water or low steam parameters	~50%	Some cases
2nd	1960–1990	Nuisance , Dust separation	Medium steam parameters	<70%	~5–14%
3rd	1990–2005	Pollution , Chemical Flue gas cleaning	40 bar/400°C	CHP	~15–22%
4th	2005–	Sustainability , Energy efficiency & Material recovery	Higher steam parameters	CHP	~25–31%

In designing WTE boilers, great care has to be paid to the layout of superheaters with regard to their position in the boiler and the flue gas temperatures. The introduction of high-temperature corrosion-resistant alloys for the cladding of the membrane walling the first pass of the boiler has allowed to operate at higher boiler wall temperatures. This also allows for increasing steam pressure and the introduction of reheating of the steam halfway the turbine. The reheating increases efficiency even with moderate superheater temperatures. The following graph relates the steam parameters to the estimated corrosion resistance of normal boiler steel (e.g., 15 Mo3) and cladding of the steel with Inconel 625 (Fig. 1).

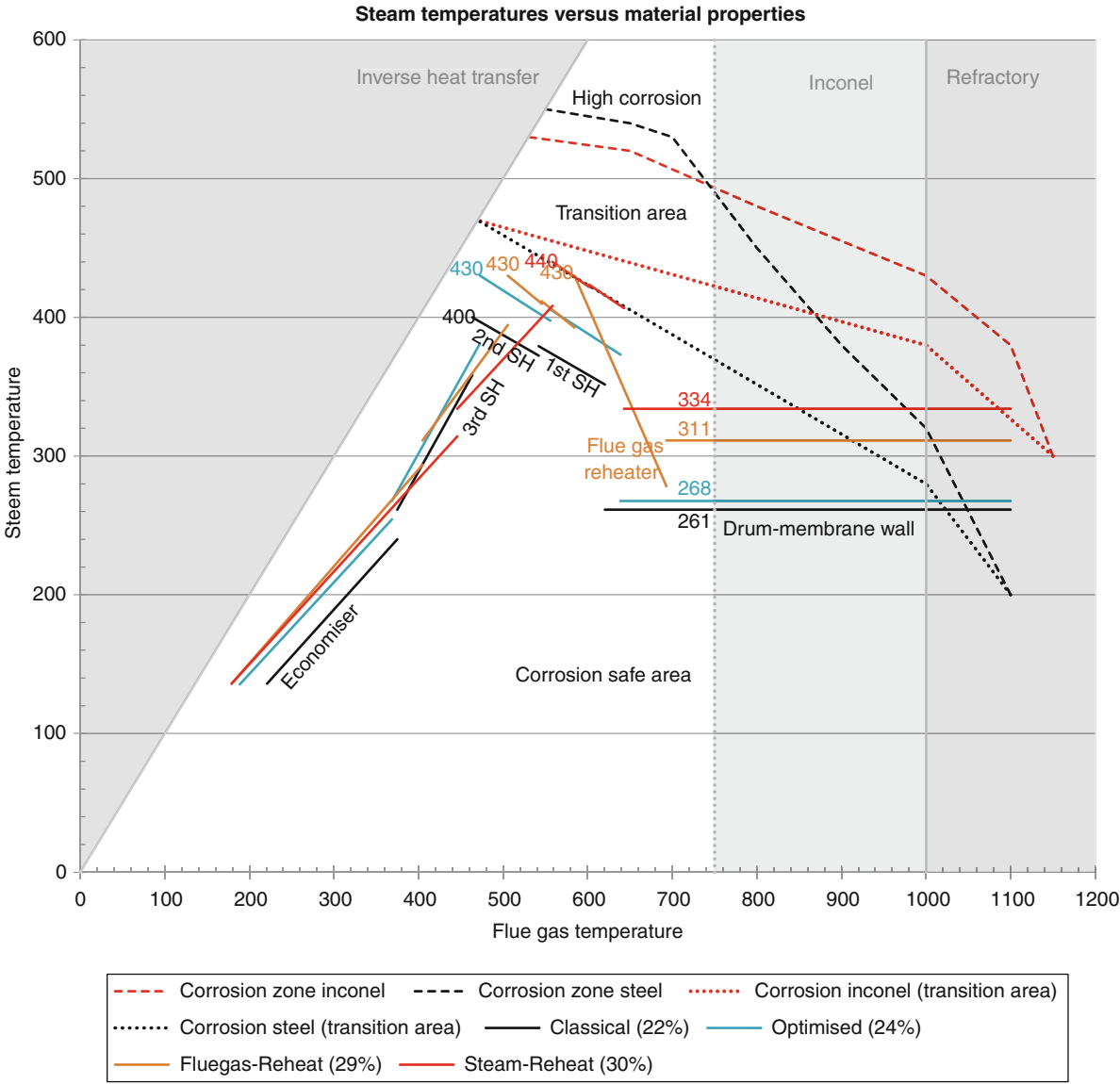
As shown in the above figure, a classical WTE boiler is operated at superheater temperatures of 40–60°C below the temperatures where corrosion starts to become important. However, it should be noted that the surface of the steel tube is between 20°C and 50°C above the temperature of the steam in the tube. The superheater temperatures for the new concepts can be increased till it is close to the transition area when the boiler design is optimized for this. Key parameters for this are especially low flue gas speeds in empty passes and in the superheater bundle. A long residence time of the flue gas allows for gradual cooling and a chemical equilibrium when reaching the superheater. Experiences with the use of Inconel in superheater bundles are mixed and are not generally cost-efficient. The good properties at high flue gas temperatures can best be exploited by cladding the membrane walls in

combination with drum pressures of 80–150 bar, thus raising drum temperature by 50–80°C. These high pressures require the use of reheating halfway through the turbine in order to prevent too high levels of steam wetness in the turbine. For this there are three rivaling concepts.

Reheating can be done by a bundle in the boiler (Rüdersdorf plant) which allows reaching high reheating temperatures (~450°C) but requires an additional bundle near the superheaters. As a consequence, the flue gas at inlet of the bundles has to be hotter, leading to more molten fly ash deposits and corrosion. This makes this a maintenance-intensive solution.

In the alternative solution, the steam is reheated halfway the turbine by a steam–steam heat exchanger that uses steam from the drum. The temperature of the reheating is limited by drum steam temperature (~330°C) but the advantage is that it is a low maintenance solution because in this case the heat transfer is shifted from bundles to membrane wall, which is much lower in maintenance due to the Inconel.

The third alternative is to use external superheating (Moerdijk, Hameln, Bilbao) to circumvent the corrosion by using heat from another fuel to superheat or reheat the steam. This is thermodynamically a very good solution combining the low-temperature heat from waste (because the corrosion limits) with the high temperature from a good fuel like Natural gas. However, the total thermal efficiency of a plant using auxiliary fuel must be divided into the efficiency expected by the fossil fuel and that achieved only



* corrosion performance of materials
are estimates for real life fluegas and flyash
conditions.
** Steam temperatures modeled in efficiency
calculation model. ©AEB-MvB

Waste-to-Energy Facilities as Power Plants. Figure 1
Corrosion: Material properties and temperatures of steam and flue gas

from the energy in the waste. Therefore, the reference efficiency for gas (Combined Cycle Gas Turbine ca. 58%) must be deducted from the total output. The resulting efficiencies from the energy in the waste have been shown to be in the same range as the steam

reheating concepts and they depend on the pressure at which the steam is delivered. Also, matching the scale of WTE boiler and the gas turbine can lead to practical problems since gas turbines are chosen preferably to be of high capacity (>200 MWe) in order to be more

efficient. Also, an economic disadvantage is that a WTE plant must be run in base load which means that the gas plant cannot be switched off during periods of low demand for electricity. This requires burning natural gas at times when electricity prices are low and makes maintenance planning critical.

Also, new WTE plants with classical steam parameters tend to have higher efficiencies due to energy optimization at other points:

- Lowering deaerator temperature from the common 135°C to 115°C
- Lowering boiler outlet temperature of 180°C further till 150°C
- After dust separation, a second economizer (for condensate preheating or for district heating) cooling the flue gasses till 90–100° (below the acid condensation point)
- A third economizer at the end of the flue gas cleaning (before induced draft fan) to extract the condensation heat of the flue gases (for condensate preheating)
- Large condenser design to improve turbine vacuum
- Flue gas recirculation as high-quantity secondary air/gas, with tertiary air above it
- Low stoichiometric combustion at 1.4 resulting in oxygen surplus of only 6%
- High efficient motor drives and the use of frequency converters to control them in order to avoid throttling by control valves
- Recovery of heat from blow-down water and steam bleeds (like deaerator off gas)
- Optimized steam bleeds from the turbine for staged heating of condensate, two-stage heating of water for district heating and primary air heating

Assessing the Energy from Waste

Efficiency Formulae

The simplest indicator of energetic efficiency of WTE power plants can be expressed as follows:

$$\text{Energy efficiency} = \frac{\text{Heat} + \text{Electricity}}{\text{Input from waste}} \quad (1)$$

In this formula, the total energy content in the waste (Lower Heating Value) is the input to the system, and the total output is the sum of the produced heat

and electricity. Many different variations have been proposed for energy efficiency. In a Dutch study alone, there are listed 11 methods that are used in different legislation ([2], p. 14). This is mainly due to different definitions, deviations from the laws of physics, and using specific parameters and assumptions.

The definitions of the energy streams can be different; e.g., the *place* where the electricity is measured can be chosen at the generator (gross electricity production) or at the point of delivery to the grid (net electricity delivery). The difference is the in-plant consumption of electricity that can amount to 2–4 percentage points in efficiency. The *time* frame can be chosen as a specific measurement, e.g., over a short time of good operation (e.g., a day), or a year, including all variations in waste, operational disturbances, shut-downs, and start-ups. The difference between short term and long term average efficiency is typically 1–2 percentage points in efficiency. Also, corrections are needed to account for fossil fuels that may be used occasionally.

The R1/D10 formula in the EU Waste Framework directive is an energy efficiency formula developed specifically for WTE:

$$\text{R1/D10 efficiency factor} = \frac{E_p - (E_f + E_i)}{0.97 \times (E_w + E_f)} \quad (2)$$

where

E_p , annual primary energy substituted by the produced heat and electricity

E_f , annual energy input to the system from fuels contributing to the production of steam

E_w , annual energy contained in the treated waste calculated using the lower net calorific value

E_i , annual energy imported excluding E_w and E_f (GJ/year)

0.97, factor chosen for accounting for energy losses due to bottom ash and radiation

This formula is now the basis of a new EU ruling that determines the distinction between considering a WTE plant as reuse (R1-status) or disposal (D10-status) plant. The threshold is $R1 = 0.6$ for plants built before 2008 and $R1 = 0.65$ for new WTE plants. The figure is not to be compared with the

energy-output-efficiency as in formula (1) shown earlier. This formula expresses how much primary fuel is replaced by the energy recovered from the waste.

$$E_p = 1.1 * \text{Heat} + 2.6 * \text{Electr} \quad (3)$$

The equivalence factors 1.1 and 2.6 are expressing that the heat produced is substituting another boiler with an efficiency of 90% and the produced electricity is substituting a power plant with an average efficiency of 38% ($1/90\% = 1.0989$ and $1/38\% = 2.6316$).

The EU R1/D10 formula for energy approach has some fundamental advantages:

- The equivalence factors account for the difference in thermodynamic quality of electricity and heat, which is a definite improvement over formula (1).
- Electrical efficiency for a plant can be measured by an established definition (ISO 1940).

This approach has some disadvantages:

- For both heat and electricity, the gross production is to be used, so internal consumption also contributes to a higher value.
- The Energy Efficiency definition is a new political definition, setting WTE power plants apart from electrical power production industry.
- Material/metal recovery is not taken into account.
- The offset from the biogenic content in the waste is not taken into account.
- The equivalence factors 1.1, 2.6, and 0.97 have no real physical background [3]. Heat is overvalued which makes it easy to fulfill the criterion in colder countries where district heating is common.
- The equivalence factor for heat does not take the temperature (and thereby enthalpy) of the heat delivery into account. This rewards (and stimulates) low-temperature district heating, but underestimates the even better effect of high-temperature steam delivery to industry. The use of *exergy* instead of energy is a possibility to physically correctly relate heat and electricity.

Many existing plants will meet the 0.6 criterion, only small plants delivering electricity only will be below. For new plants to be built, the 0.65 is not challenging for full optimization. Values up to the range of 0.9 till 1.1 have been reported.

Alternative Efficiency Evaluations

In order to address the performance of a WTE power plant in a specific context, several more integrated methods are available:

- Instead of efficiency (which is a ratio), the direct output can be expressed. Classical plants deliver about 500–650 kWh/t of waste when this has the average calorific value of 10 MJ/kg (standard waste, Lower Heating Value LHV). For the high-efficiency plants, this can be raised to as much as 850 kWh/t of waste. For presorted high calorific wastes with a calorific value of 12–14 MJ/kg, the recovery can reach 900–1,100 kWh/t of waste. Additionally, 1,500–2,500 kWh-heat/t of standard waste can be delivered with only a small de-rating of the capacity to produce electricity. De-rating depends on the temperature at which the heat is delivered and the configuration of the turbine, and can be 13–20% of the amount of heat extracted from the turbine, i.e., delivery of 1 kWh of heat reduces electrical output with 0.13–0.20 kWh.
- The use of *exergy* instead of energy is a possibility to physically and thermodynamically correctly relate heat and electricity, see [4, 5]. Exergy is the equivalent amount of work that can be performed by a stream of heat. A small formula uses the temperature to convert heat energy in it to equivalent exergy. This solves the use of arbitrary equivalence factors as used in the R1/D10 formula and expresses the output of a CHP plant in an equivalent amount of electricity that that plant could produce if no heat was extracted from the turbine.
- CO₂-evaluation is calculating the overall impact of the greenhouse gas effects. This constitutes the direct emission of CO₂ which originates partly from fossil origin (~450 kg/t of waste). The other half of the carbon in the waste is of biogenic origin content of the waste which is CO₂-neutral. The direct emissions are offset by the external effect of substitution of other fossil energy power production by the energy produced from waste. Also, the substitution of mining and materials processing by the recovery of metals can be calculated. For normal WTE installations, the combination of biogenic content and substitution results in a net greenhouse

effect of +100 to about zero kg CO₂/t of waste. For high-efficiency WTE installations, the higher substitution leads to a net avoidance ($\sim$ negative emission) of up to -200 kg CO₂/t of waste [6].

- Life Cycle Analyses (LCA) is doing the same as a CO₂-evaluation but it also adds a list of other impacts on the environment and health, like: human toxicity, ecotoxicity, acidification, eutrophication, injuries, mineral extraction, land use, nonrenewable energy, and ozone layer depletion. In the LCAs for different methods to manage wastes, it is generally shown that energy efficiency has a major effect in reducing the impact of waste management options. The process to conduct these studies is much more complicated but allows detailed analyses of the impacts [7].
- Cost Benefit Analyses (CBA) is doing the same as an LCA but monetizing all the effects. So comparisons of alternatives in a specific situation can be done using a single economic value. The economic value of different impacts has however high uncertainties which make the use of CBA often complicated and strongly depending on the boundary definitions of the system to be considered.

Future Directions

In the EU, there are strong legal drivers (e.g., formula for R1/D10 distinction and Best Available Technology Reference (BAT-BREF) documents) and in many nation members economic incentives (sustainable energy premiums) that encourage striving for higher efficiency of WTE. This promotes the use of more of the available energy efficiency optimization options. Even with classical steam parameters, new plants tend to be more energy efficient. WTE plants that produce only heat are becoming rare (with the exception of steam delivery to industry) and even in the case of Combined Heat and Power (CHP), electrical output is optimized. In the U.S.A., there is renewed interest in WTE which could also result in new high-efficiency plants. In Asia, a first wave of conventional installations is being built in China while in Japan there is up to now more focus on ash treatment than on energy efficiency.

It is to be expected that Europe will keep its technological leading position in getting more energy out of waste. After new advances in technology have been proven in the E.U., it is expected that they will also be implemented elsewhere.

As LCA studies have shown, consistently, that energy efficiency is a key parameter for reducing the overall cost as well as the environmental impact of waste management, it is expected that minimum standards for energy efficiency will be raised over time. Economical modeling shows that the additional energy produced by WTE power plants is among the least-cost sources of sustainable, and also renewable, energy.

Bibliography

1. Berlo MAJ, van en Wandschneider Jörn (2003) High efficiency waste-to-energy concept [Sectie van boek]//Municipal solid waste management, strategies and technologies for sustainable solutions/boekaut. Christian Ludwig, Hellweg Stefanie en Stucki Samuel. Springer, Berlin, Heidelberg. ISBN 3-540-44100-X
2. Erbrink H et al. (2001) Analysis of energy efficiency definitions for policy making (in Dutch) [Boek]. NOVEM/KEMA, Utrecht, publicatiecentrum@novem.nl. - Reportnr: 2EWAB01.03
3. Tebert C (2006) The energy efficiency formula of annex ii of the waste framework directive: a critical review. [Rapport] = R1/D10 formula Review/Ökopol gmbh, Institute for environmental strategies – [sl] : Ökopol. <http://www.eeb.org/activities/waste/20060630-Okopol-Brief-on-MSWI-efficiency-formula-v5-final.pdf>
4. Murer MJ et al. (2009) Comparison of energy efficiency indicators for energy-from-waste plants [Conferentie]/International waste management and landfill symposium, Sardinia: [sn], vol 12
5. Murer MJ et al. (2010) Exergetic analysis of heat transfer and efficiency in Efw plants [Conferentie]/Third international symposium on energy from biomass and waste. CISA, Environmental Sanitary Engineering Centre, Venice, Italy, vol 2010
6. Berlo MAJ, van en Waart Harry de (2008) Unleashing the power in waste, Comparison of greenhouse gas and other performance indicators for waste-to-energy concepts and landfilling [Conferentie] = Unleashing the power in waste// North American waste-to-energy conference. Philadelphia, Pennsylvania, USA: NAWTEC, 2008. vol 16, p 16, NAWTEC16-1937
7. EU – JRC (2007) Environmental assessment of municipal waste management scenarios: Part II – Detailed life cycle assessments [Rapport] = JRC 41241/EUR 23021 EN/2/Institute for environment and sustainability. European Commission, Joint Research Centre, Ispra, Italy. p 182 - <http://www.jrc.ec.europa.eu>. ISBN 978-92-79-07450-9

Waste-to-Energy for District Heating

LASSE TOBIASEN, BETTINA KAMUK

Ramboll Energy – Waste-to-Energy, København S, Denmark

Article Outline

Glossary

Definition of the Subject

Introduction

Waste Boilers

Grate-Fired Boilers

Different Types of Boiler Arrangements

Steam Turbines for Electricity Production

Process Concepts for a Complete Electricity-Only

Energy for Waste Facility

WTE Facilities Delivering “Small-To-Medium”

Amounts of Heat

Larger District Heating Networks and District Heat-

Producing WTE Facilities

Advanced Solutions: Heat Pumps and District Cooling

Optimizing Energy for Waste Facilities for CHP

Operation

Conclusions and Future Directions

Bibliography

Glossary

District heating (DH) A network of pipes, typically underground, that distributes hot water (or steam) used for space heating of residential homes and industrial spaces. This network allows for centralized production of heat to be distributed to many consumers.

Waste-to-energy (WTE) A thermal process where the chemical energy contained in solid wastes is converted to electricity and/or heat.

Definition of the Subject

In Denmark, waste incineration has been used to provide district heating since 1903, when the Frederiksberg Municipality in Copenhagen realized that landfilling of municipal waste could not continue within the municipality boundaries [1]. The incineration plant of the municipality provided steam, hot water, and

some electricity to a nearby hospital, while reducing the waste volume and mass at the same time.

The first waste-to-energy (WTE) plants in Denmark were producing heat only, while from the late 1980s, they started to produce combined heat and power (CHP). However, recovery of district heat for useful purposes requires large investments in energy infrastructure (district heating pipes distribution and transmission network), and the lack of such infrastructure makes combined heat and power difficult in many countries. Most countries at this time produce only electricity (and not heat) from their waste incineration plants.

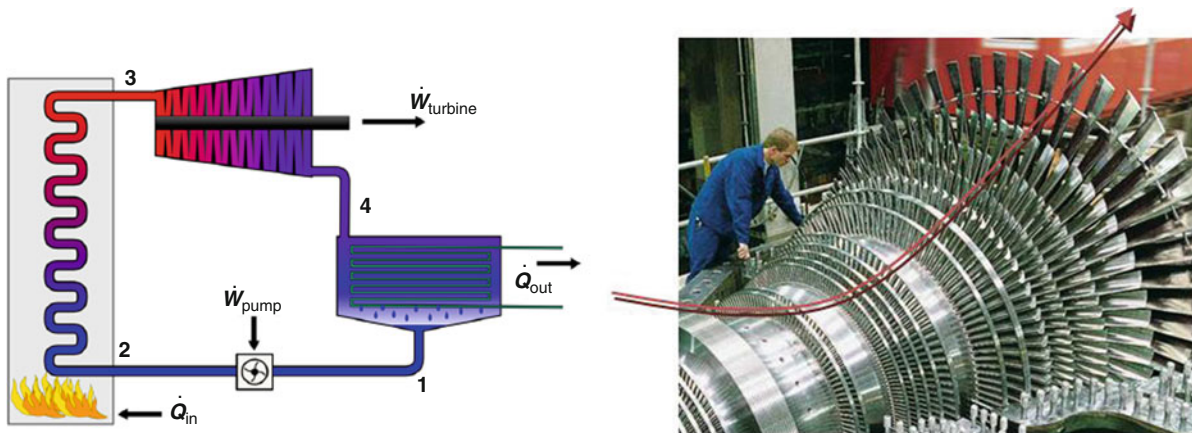
Nevertheless energy recovery in district heating waste-to-energy plants is three- to fivefold higher than from electricity-only plants. Additionally, the revenues for a waste plant from selling district heat may substantially reduce the waste gate fee. For example, in 2007, 70% of the costs of the 29 Danish WTE plants were recovered by revenues from selling district heat. In Denmark today, about 20% of the total district heat production and approximately 4.5% of the electricity used are due to energy recovery from waste incineration [2].

This entry reviews waste-to-energy and district heating technologies, drawing on the considerable experience accumulated in Denmark.

Introduction

This entry deals with energy recovery in waste-fired boilers, using the Rankine steam cycle for electricity production; and also examines methods of producing district heat or process steam that allow combined heat and power production. The Rankine thermal cycle is used to produce approximately 90% of the world's electricity.

Figure 1 illustrates this process in its simplest form: Water is pressured in a pump, superheated in a boiler, and expanded in a steam turbine. It can be shown thermodynamically that the electricity yield at the turbine, $\dot{W}_{\text{turbine}}$, is much higher than the power consumed by the pump, $\dot{W}_{\text{pump}}$. It can be shown that the cycle efficiency – from fuel to turbine output – increases with higher steam temperatures and pressures (point 3); similarly, a higher efficiency is reached by lowering the steam condensing temperature (point 4). The



Waste-to-Energy for District Heating. Figure 1

Illustration of the Rankine water-steam cycle for electricity generation, as well as picture of a steam turbine rotor (Figure to the left from wikipedia.org. Picture of steam turbine rotor by Courtesy of MAN Diesel Turbo SE, see also <http://www.mandieselturbo.com>)

condensing temperature ultimately depends on the cooling temperature level available to the condenser; the condenser can be either water cooled (seawater or river/lake water) or air cooled. This is why power stations located in cold regions (access to cold seawater, for instance) have a higher electrical efficiency than power stations without such access.

Waste-to-energy facilities in Europe are often equipped with air-cooled condensers, which can be either wet (water consuming) or dry, and configured with either forced airflow or natural draught (“cooling tower”). Air-cooled condensers are the most common choice of WTE facilities.

Figure 2 shows the principle of such an air-cooled condenser system. In the top illustration in Fig. 2, low-pressure high-volume steam from a steam turbine exhaust is routed to a four-row air-cooled condenser unit, where the steam is condensed in finned tube bundles; large fans (often 10 m in diameter) force airflow across the condenser heat surface areas, thus removing the heat. Air-cooled condenser installations are necessary where there are no other means of cooling.

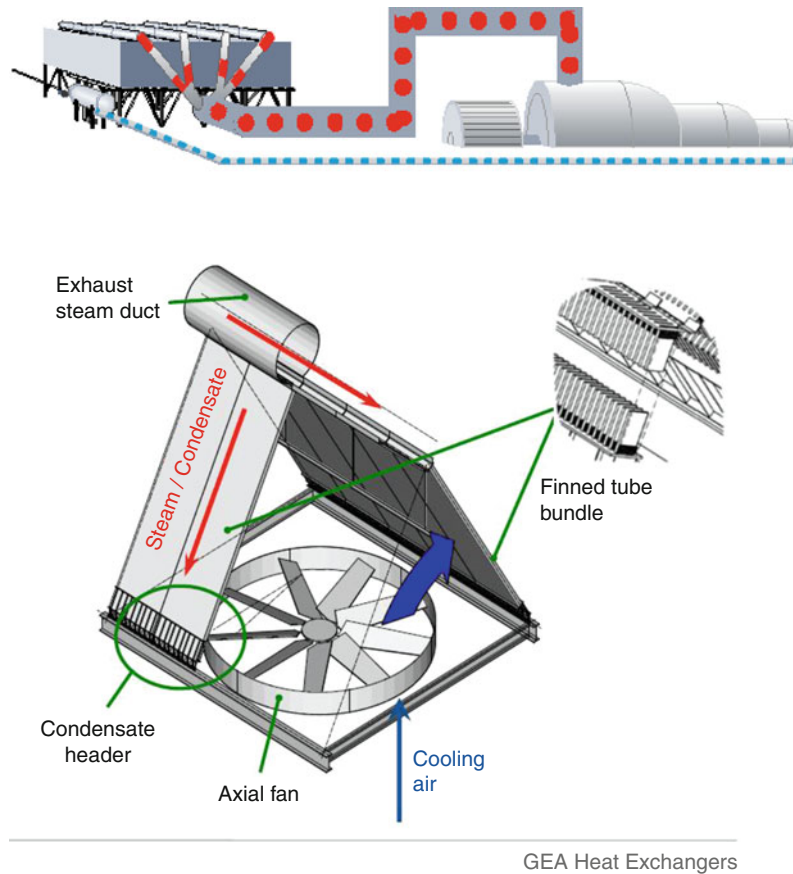
Therefore, WTE facilities with no access to a heat user must reject large amounts of heat through condensers, such as shown in Fig. 2. Typically, the net electrical efficiency of WTE facilities is in the region of 25%. Hence, for a 100 MW plant (corresponding

to 32 t/h of waste with a calorific value of about 11 MJ/kg), roughly 25 MW of electricity can be sold to the grid. Of the 100 MW energy input, about 55 MW is rejected in water- or air-cooled condenser. Therefore, a very large amount of energy is lost.

However, if there is a demand for heat, such as space heating, in the vicinity of the WTE facility, it may very well be feasible to provide heat from the waste facility – thereby utilizing part of the heat rejected to condensers. Other options for using the WTE heat include the supply of process steam to a nearby factory or industry that would otherwise be produced by oil or gas.

Fuel input (waste):	100%
Electricity from waste:	~25%
Heat reject in Air-cooled condenser:	~55%
Other losses:	~20%
Useful heat potential (hot water or steam):	~55–65% (Flue gas condensation, a method of extracting additionally 10–15% of the fuel input as useful heat, but this is not included in this table)

There are various methods of producing heat from a WTE plant, as discussed in the following sections.



Waste-to-Energy for District Heating. Figure 2

Air-cooled condensers (Figures from <http://www.gea-energietechnik.com>, Courtesy GEA Energietechnik)

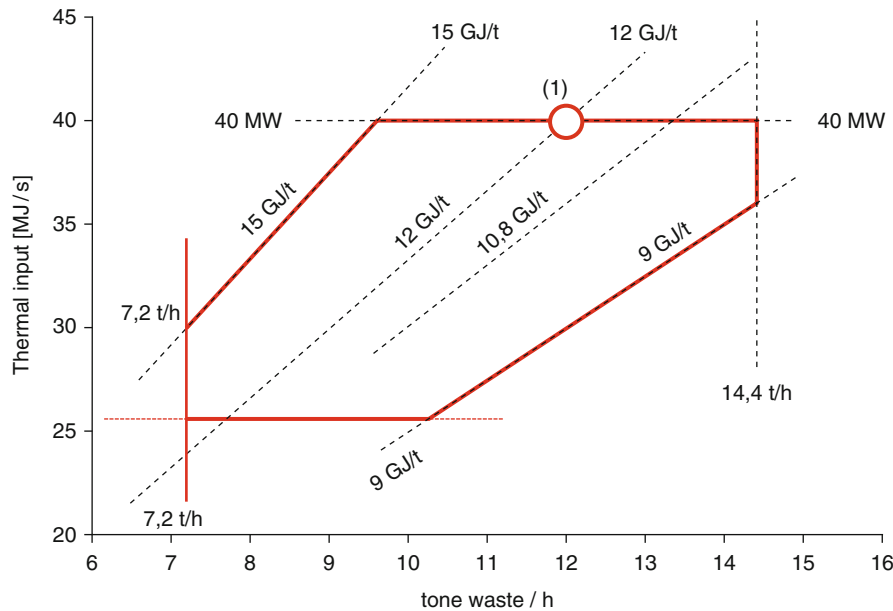
Waste Boilers

Waste incineration plants are designed to treat waste with great variation in the composition of the incoming waste (This chapter is mainly based on the following source: solid waste technology management, vol 1, chapter 8.1, as well as generally based on Ramboll knowledge and experience). This is the primary difference between waste incineration and other combustion systems, and it has large implications on the design of the incineration plant. Waste is not an ideal fuel and it is important to link the fuel characteristics to the combustion process. The main technologies described in this section are combustion of

as-received municipal solid wastes (MSW) by means of a moving grate.

Common to waste boilers is the concept of the “capacity diagram,” which determines limits of thermal capacity (energy input) for a particular boiler, given the variation of heating value in the waste. In the example shown in Fig. 3, the boiler can accept waste with a (lower) heating value from 9 to 15 MJ/kg, and a throughput from 7 to 14.4 t/h of waste.

Typical heating values for some waste fractions are as shown below [3]. Note the very large variations, and hence the need for a waste boiler to handle fuel input with a variety of lower heating values.



Waste-to-Energy for District Heating. Figure 3

Typical waste boiler capacity diagram. The nominal load point in this example is indicated by point (1), where the waste has a heating value of 12 MJ/kg

Fresh wood (50% moisture)	10 MJ/kg
Oven dried wood	17–19 MJ/kg
Paper/cardboard	14–15 MJ/kg
Plastic	28–35 MJ/kg
Metal, glass, concrete	0 MJ/kg
Organic waste	4 MJ/kg
Municipal waste (Copenhagen)	10–11 MJ/kg
Municipal waste (Bangkok)	2.5–12 MJ/kg

The minimum thermal load of a modern grate combustion waste boiler is 65–70%, limited by the ability to ensure proper burnout and upholding legislative requirements with regard to emission limits (like CO, Dioxins, etc.) and flue gas temperature limits (flue gas at least 850°C for a minimum of 2 s, for instance).

Grate-Fired Boilers

A combustion grate is a transport device that moves the burning fuel from the inlet through the furnace to the

bottom ash outlet. During transportation, the fuel is mixed and combustion air is added. Volatile material is released to the combustion chamber above the grate and fixed carbon is burned on the grate. The grate is an integrated part of the furnace, where the fuel is converted into energy. The grate ensures drying, ignition, combustion and energy release, and complete burnout before the bottom ash outlet [4].

Flue gases leaving the grate enter the actual “boiler,” where the main steam generation takes place. A modern European waste boiler often consists of three empty flue gas passes, in which the flue gas is cooled to approximately 600–750°C while providing heat to evaporate the boiler water to generate (saturated) steam. The dominant means of heat transfer in this part of the boiler is radiation, due to the high temperature of the flue gases; thus, the first three passes are often called “radiation passes.”

Boilers for waste facilities are “drum-type” boilers, characterized by a “drum” maintained below the critical steam pressure of 221 bar. The large cylindrical drum contains water at high pressure and is located at

the top of the boiler. Water circulates by natural convection by means of “down-comer” tubes and up through the evaporator tubes, since evaporated water (steam) has a lower density than the water in the down comers. Typically, the circulation ratio is 15–20, i.e., the circulating mass flow in the evaporator section is 15–20 times the dry steam flow exiting the drum.

A very simple analogy to the circulation principle in a drum-type boiler is a household electric drip brewing coffeemaker, where hot water is also moved upward without need of a pump, by the thermosyphon method.

When the flue gases enter the fourth pass of the boiler, the primary method of heat transfer changes to convection: Superheater tube bundles are located in the path of the flue gas flow, thus providing a large surface area to facilitate the transfer of heat from flue gas to steam. The superheater tube bundles heat the saturated steam from the drum to the desired final steam temperature.

The economizer is the last set of tube bundles in the boiler, seen from the flue gas point of view. Here the inlet water is heated to somewhere below the boiling point at the particular water pressure (e.g., heated to

200°C) while the flue gas is cooled to the desired set point temperature, e.g., 180°C.

Typically the pressure in the drum is 50–70 bar, and the final steam temperature 380–440°C.

The mechanisms described below are illustrated in Fig. 4.

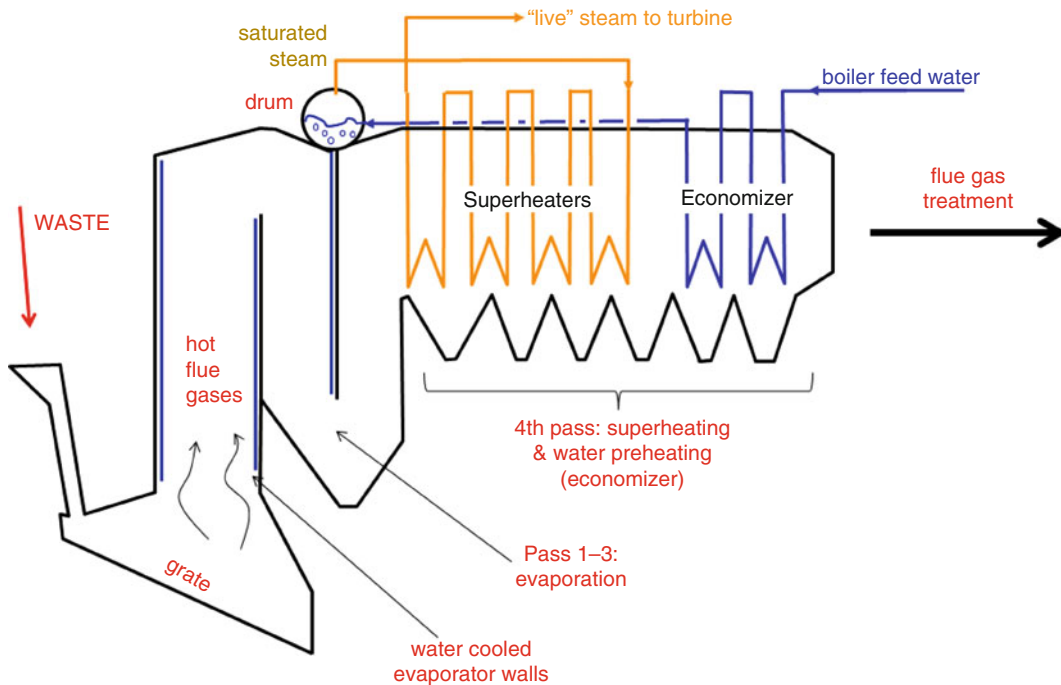
Different Types of Boiler Arrangements

Depending on the needs and requirements of the purchaser of a WTE facility, the boiler heat transfer surfaces can be arranged in various ways, thus influencing up-front cost, operating and maintenance costs, availability, and efficiency.

The most common boiler configurations are described below (Table 3):

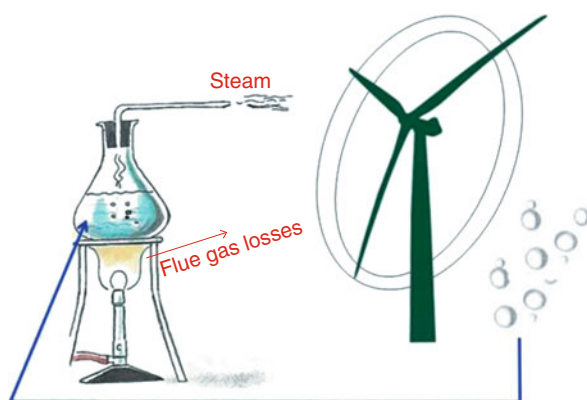
Steam Turbines for Electricity Production

The steam turbine is the equipment that transforms the energy in the steam (temperature and pressure) to mechanical energy, via a rotating generator that produces electricity. The thermodynamic process is



Waste-to-Energy for District Heating. Figure 4

Schematic of a modern European 4-pass waste boiler



Waste-to-Energy for District Heating. Figure 5

Simple illustration of the steam cycle, with a wind-turbine analogy

expansion of the high-pressure steam that produces kinetic energy (high velocity) that is transferred to the blades of the steam and forces it to turn through an electromagnetic field.

This process, in very simple terms, can be compared to a wind turbine with the airflow past the wings (or “blades” as it is called in a steam turbine) being replaced by hot steam under pressure (Fig. 5).

It is often useful to plot a steam turbine expansion in a so-called mollier or h s diagram. “ h ” is the steam enthalpy and “ s ” is the entropy, both of which are thermodynamic variables of state, and can be thought of, as variables that can be looked up in tables provided that temperature and pressure are known (in the wet steam region this is not enough – here also steam dryness must be known).

In such a diagram, the enthalpy change during the steam expansion is shown “visually,” which has particular interest due to the first law of thermodynamics stating the relationship between mechanical work, heat loss, and enthalpy drop (see equation below; its validity is subject to various assumptions for steam turbines, such as steady-state operation and negligible heat loss).

$$W[\text{Work}] = m \times (h_{\text{in}} - h_{\text{out}})$$

$$[\text{kW}] = [\text{kg/s}] \times [\text{kJ/kg}]$$

Since the heat loss from a steam turbine is negligible, and considering that approximately 97–99% of the mechanical shaft work can be translated into electrical power, the above equation shows that the steam turbine power output is mainly dependent on:

1. Mass flow of steam (given by the size of the boiler)
2. Inlet and outlet enthalpy (given by steam parameters, and available cooling, which sets the final expansion pressure)

A steam expansion from 40 bar to 1 bar is shown in Fig. 6 below, by means of the red line between the two yellow dots. The enthalpy drop on the y-axis can be read to be $(3.24 - 2.56) \text{ MJ/kg}$ 0.68 MJ/kg. Therefore, for every kg/s of steam subjected to such an expansion, roughly 0.68 MW of electricity can be produced.

The “internal” efficiency of a steam turbine can be determined by the steepness of the red expansion line, where a completely vertical line represents 100% efficiency (limited by the second law of thermodynamics). This would correspond to an expansion to the green dot.

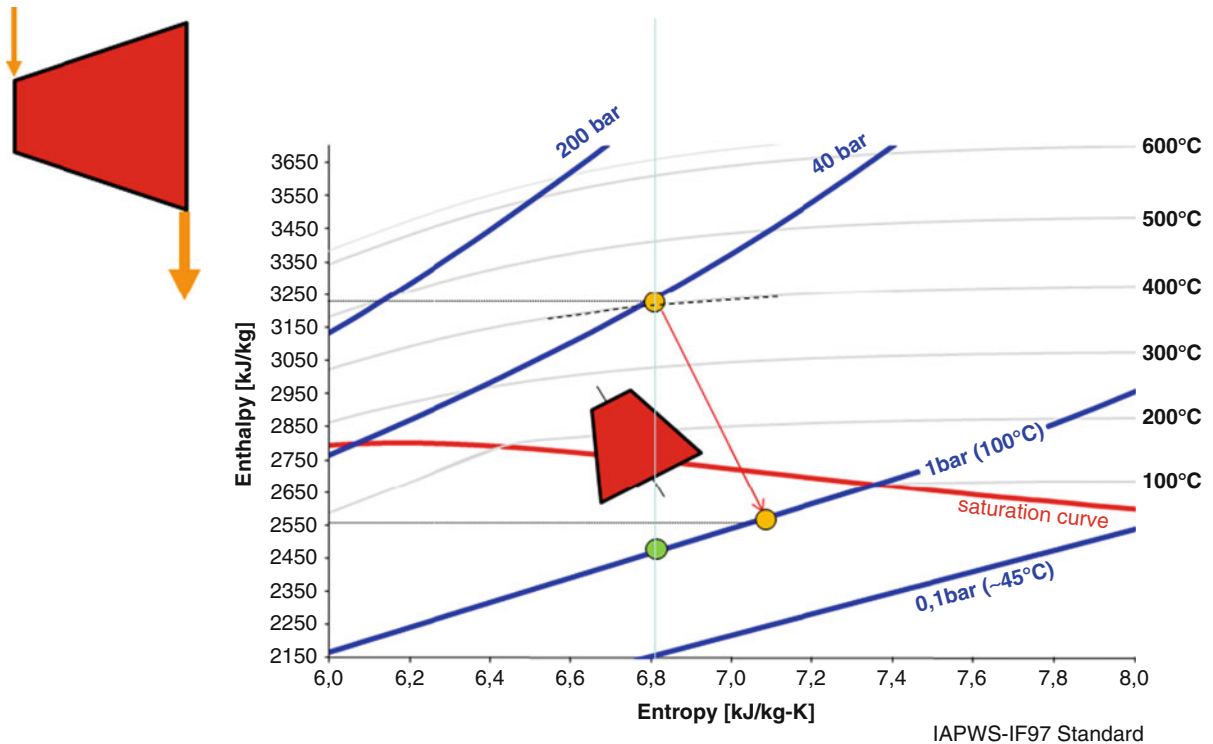
Figure 6 also clearly shows that each kg/s of steam would produce additional electricity if the inlet pressure and temperature were higher (see for instance the “200 bar” line) and if the outlet pressure were lower (see the “0.1 bar” line).

As noted earlier, thermal cycles are more efficient at higher steam temperatures and pressures. However, due to high-temperature corrosion of boiler tubes arising from chloride and other substances in the waste, waste boilers are for practical purposes limited to approximately 440–450°C and 60–90 bar, depending on the levels of maintenance a plant owner will accept.

Other examples of expansion lines can be seen in Fig. 7, using 8 kg/s as an example. This figure shows that, at 90% steam turbine efficiency, the potential power output is 4.1 MW, for an inlet state of 400°C and 40 bar and a fixed back pressure of 2 bar. This drops to 3.2 MW for an efficiency of 70%, and drops further to 2.6 MW when the steam temperature and pressure are decreased to 380°C and 22 bar.

Process Concepts for a Complete Electricity-Only Energy for Waste Facility

When the waste boiler and steam turbine technologies are combined, along with appropriate flue gas treatment, a waste-fired power plant concept is formed. Such a concept is illustrated in Fig. 8, comprising a horizontal-type boiler with a vertical economizer, a 400°C/40 bar steam turbine connected to an air-cooled condenser, and a dry flue gas cleaning system. The thermal cycle also



Waste-to-Energy for District Heating. Figure 6

h-s diagram for a hypothetical steam turbine expansion (Source: Ramboll)

shows other necessary components like a deaerator, which removes non-condensable gases from the boiler water/condensate. The process flow path of the water-steam flow and the flue gas, respectively, is shown in Fig. 9.

Experience with waste boilers has shown that the risk of high-temperature corrosion of the superheater tubes is greatly dependant on the flue gas temperature at the entrance to the fourth pass. It is therefore often a boiler design criterion that the flue gas does not exceed roughly 625–650°C, when passing across the hottest superheater tubes. Selection of correct/appropriate material (steel alloy) for the tubes is also important, although developments in anti-corrosive steel do not appear to have solved the corrosion problems in waste boilers.

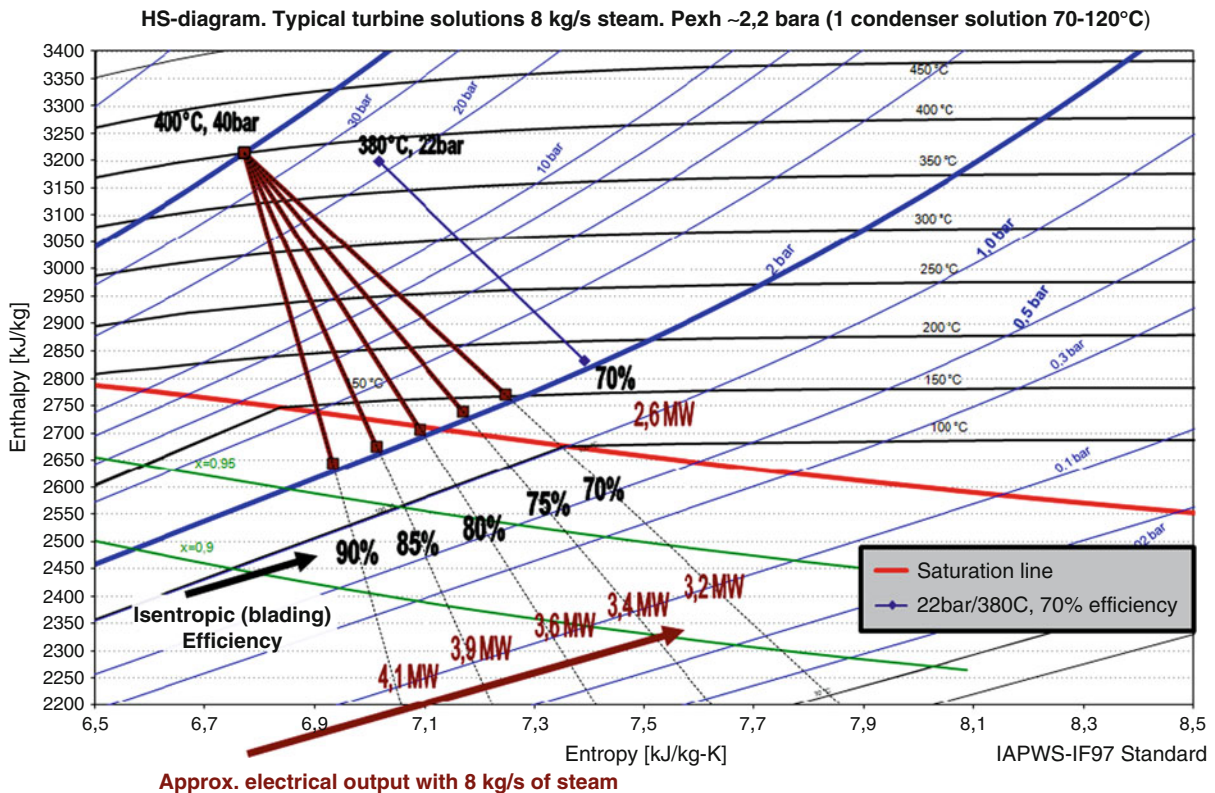
Superheater tubes producing 400°C steam on a municipal waste facility are estimated by Ramboll to have a lifetime of up to 10–15 years, under the best conditions and with a modern boiler design. Even moderate increases to this temperature,

say up to 450°C, may cause severe corrosion and reduce the lifetime of superheater tubes to perhaps 3–5 years. Ramboll estimates that further increases, say to 500°C, may entail boiler tubes maintenance several times a year due to severe high-temperature corrosion.

Waste boiler manufactures generally do not recommend higher steam temperatures than 440–450°C, unless special circumstances dictate otherwise. Methods of reducing corrosion include choosing high alloy steels, or specially resistant coatings.

WTE Facilities Delivering “Small-To-Medium” Amounts of Heat

An “electricity-only” waste-to-energy plant can easily be designed to produce small-to-medium amounts of heat; for example, when heat or process steam can be sold a number of months of the year (winter) while the remaining months, the plant can operate to produce electricity only.



Waste-to-Energy for District Heating. Figure 7

Various expansion lines for 8 kg/s steam. The 2 bar back pressure equates to the necessary pressure level to produce hot water at 120°C (Diagram by Ramboll)

A technical solution to this is to have an “extraction” turbine, with a crossover valve placed somewhere along the turbine expansion line. When actuated toward the close position, the valve forces out steam to a hot water condenser which (Fig. 10) produces 80–110°C water. The crossover valve can thus divert steam to where it is mostly needed: either to produce heat, or to produce electricity, as required. As such, this type of turbine enables both “heat and power” operation as well as “electricity-only” operation.

Turbines such as the one shown in Fig. 10 could be the choice when the plant is connected to a small district heating network, or for the plant to be “CHP ready,” for future district heating operation.

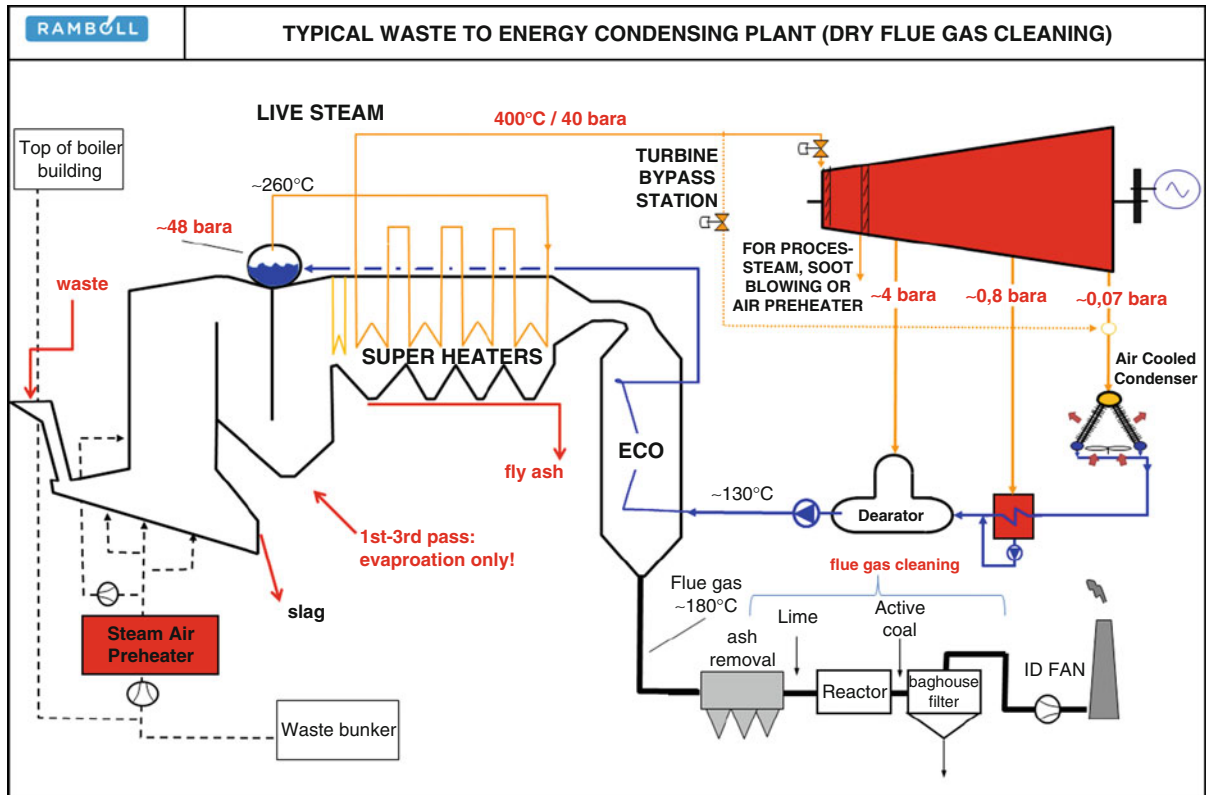
Figure 11 shows a cross-sectional view of an “extraction” steam turbine with the ability to operate freely between “electricity-only” and “combined heat and power” modes.

Figure 12 is an elaboration of Fig. 10, now showing also the major control principles relevant for an extraction turbine connected to a small district heating network.

The waterside district heating pumps (shown as a set of three pumps, for redundancy reasons) control to achieve a desired pressure drop in the district heating network, dP , for instance 2 bar. This allows enough pressure to operate the heating installations connected to the network.

If the heating consumers in the network turn on their radiators (call for more heat), the pressure in the network will drop, and the district heat pumps of the plant will respond by increasing the flow to again uphold the desired dP .

The steam turbine crossover valve, as well as the steam valve in the line to the DH condenser, is regulated to keep a constant district heating supply temperature. Thus, if the consumer calls for more



Waste-to-Energy for District Heating. Figure 8

Complete water-steam cycle for a typical European waste-to-energy facility, with no heat production (Source: Ramboll)

heat, the district heating pumps will increase the flow and the initial consequence will be a lower DH temperature. This is picked up by the turbine controls, which respond by providing more steam to the condenser until the set-point district heat temperature is obtained.

These controls in combination will thus match the consumer demand for heat (in MJ/s) and desired supply temperature, which is a set point determined by the heating network operators. For large transmission heat networks, the supply temperature is 100–120°C or more, whereas smaller heat networks may have supply temperatures as low as 75–85°C.

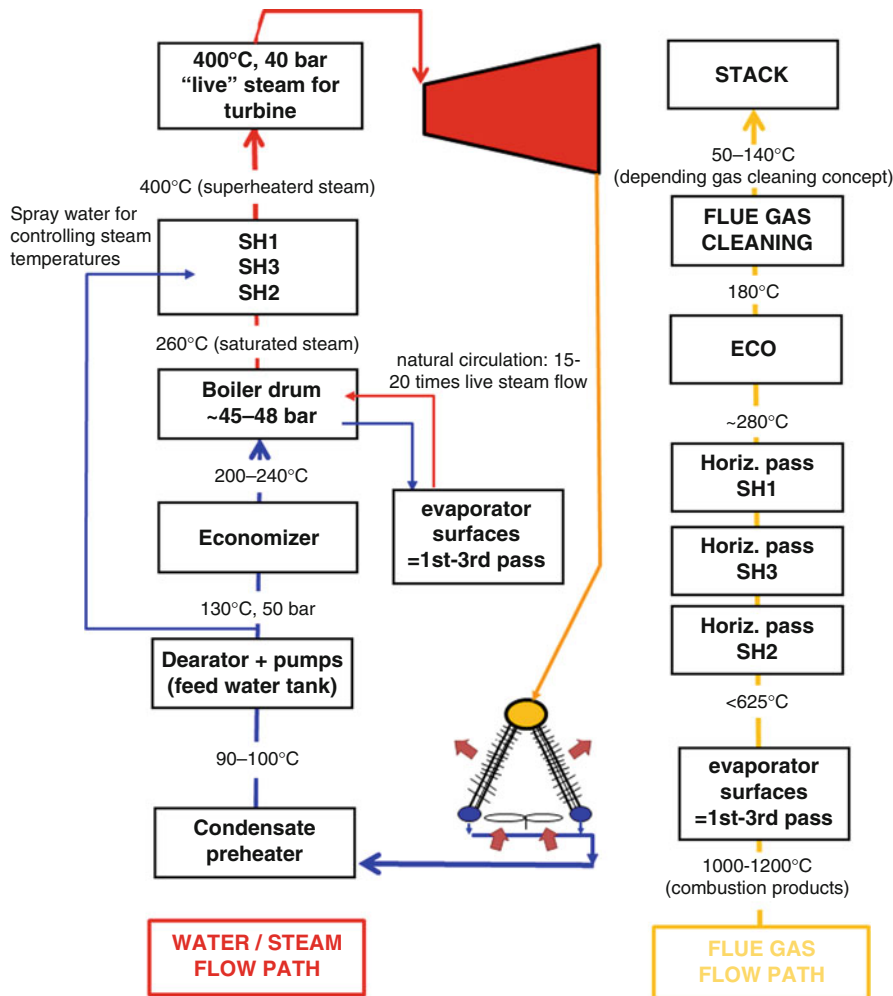
Generally, the supply temperature is a function of the ambient temperature, since a higher supply temperature will allow the transfer of more heat in the district heating pipework, which is also bound by hydraulic limits.

Table 1 below shows an example of set points for district heat network supply temperatures, as a function of ambient temperature.

Larger District Heating Networks and District Heat-Producing WTE Facilities

When a WTE facility has access to a large district heating network, such as in countries with a cold climate like Denmark, Sweden, and Finland, it is common to establish so-called back pressure turbines, where all the steam exiting the turbine is used to produce district heat (Fig. 13).

This configuration is illustrated in Fig. 14 (next page) where the very large air-cooled condenser has been replaced by two, and much smaller, surface steam condensers, where the condensing heat is transferred to the district heating water. To exemplify other options, a 14-bar process



Waste-to-Energy for District Heating. Figure 9

Process flow path for water-steam (left) and flue gas (right) in a WTE facility (Source: Ramboll)

steam extraction to a hypothetical nearby factory is also shown.

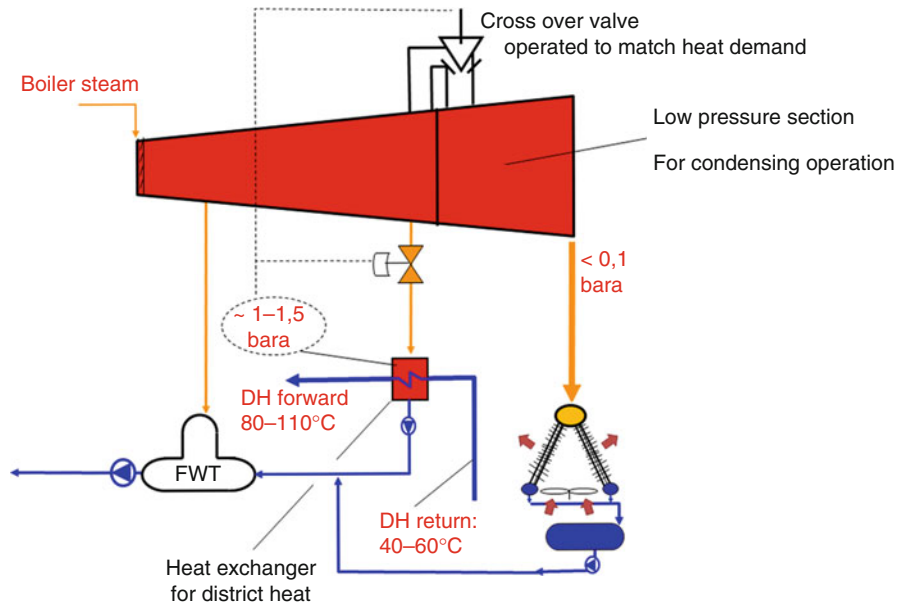
Table 2 compares the electrical and total efficiencies of an “electrical-only” and combined heat and power “district heating” plant, configured as discussed in this chapter. Since the district heating water induces a higher back pressure (~ 1 bar) than an air-cooled condenser (~ 0.1 bar), the electrical efficiency of the “electrical-only” plant is higher.

As a result, the electrical-only plant will have an electrical efficiency approximately 3–6% points higher than the corresponding district heating plant (based on various assumptions).

Table 2 below shows typical figures based on a 100 MW thermal plant as an example. It is clear that the district heating plant produces 3.5 MW less electricity.

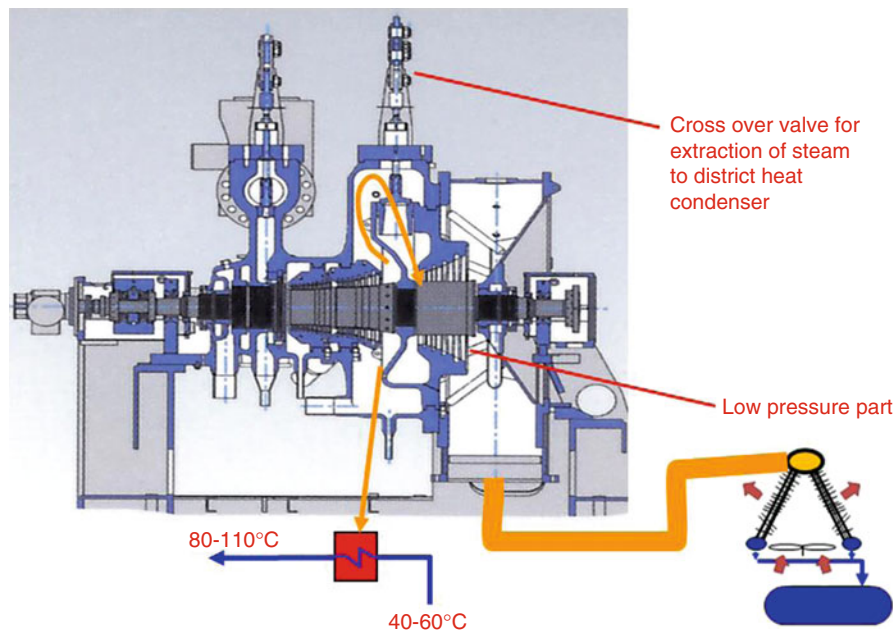
However, a total of 63 MJ/s of heat is produced instead – hence 1 MW of electricity is converted to 18 MJ/s of heat ($63/3.5 = 18$). This is 18 times more efficient than electric space heating, and also even more efficient than highly efficient heat pump installations, which can produce up to approximately 4–6 MJ/s of heat for every MW of electricity.

The conclusion is that a waste-to-energy district heating turbine is an extremely efficient means of producing heat. Obviously, the “value” of the produced



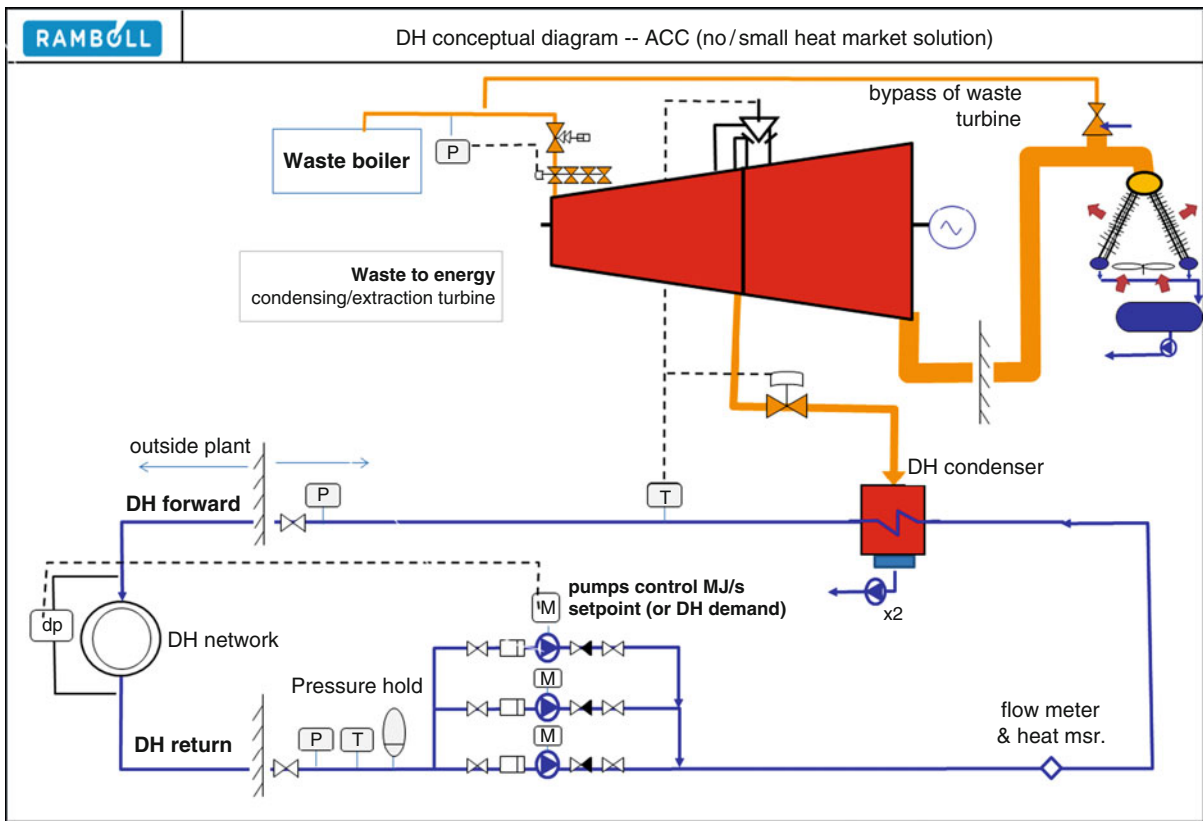
Waste-to-Energy for District Heating. Figure 10

Extraction turbine with the ability to provide variable quantities of hot water, depending on the heat requirement. The depicted solution is applicable for a plant to be operated both in "electricity-only" mode, as well as in "combined heat and power" mode (Figure by Ramboll)



Waste-to-Energy for District Heating. Figure 11

Extraction steam turbine cross section. Crossover valve seen used to drive steam to the district heating condensers as per required heat demand (Courtesy of MAN Diesel & Turbo SE, see also <http://mandieselturbo.com/0001157/PrimeServ/Turbomachinery/Steam-Turbines.html>) (Figure by courtesy of MAN Diesel and Turbo SE, annotated by Ramboll)



Waste-to-Energy for District Heating. Figure 12

Typical districting heating components and principle control mechanisms

Waste-to-Energy for District Heating. Table 1 Example of DH supply temperature versus ambient temperature for large Swedish DH network

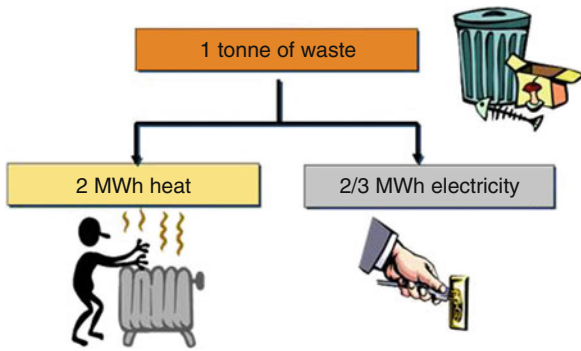
	T ambient (°C)	T supply (°C)	T return (°C)
Very cold winter	<−15	115	52
Cold winter	−10 to −15	105	50
Winter	0 to −10	95	45
Spring/autumn	+10 to 0	85	42
Summer	> +10	75	45

heat depends on the local heat market and the size of the district heating network. Figure 15 shows the district heating network in the Copenhagen region that stretches over a distance of about 40 km from east to west.

The high total efficiencies of waste-to-energy plants connected to (medium-large) district heating networks result in low incinerator gate fees, since there is generally a higher energy-sales income than electricity-only plants.

The configuration of pumps and main control strategies for a district heating WTE facility can be seen in Fig. 16. The mechanisms are slightly different than the “electricity-only” concept shown in Fig. 12, since now the district heating pumps are controlled to sustain a given district heating supply temperature. The result is that the district heating pumps ensure that variations in the steam flow are taken up by the district heating water flow.

If, for instance, the thermal load on the waste boiler was to decrease (say due to a lower heating value of the waste), the steam flow to the turbine would be lower. The immediate consequence would be a lower district heat supply temperature.



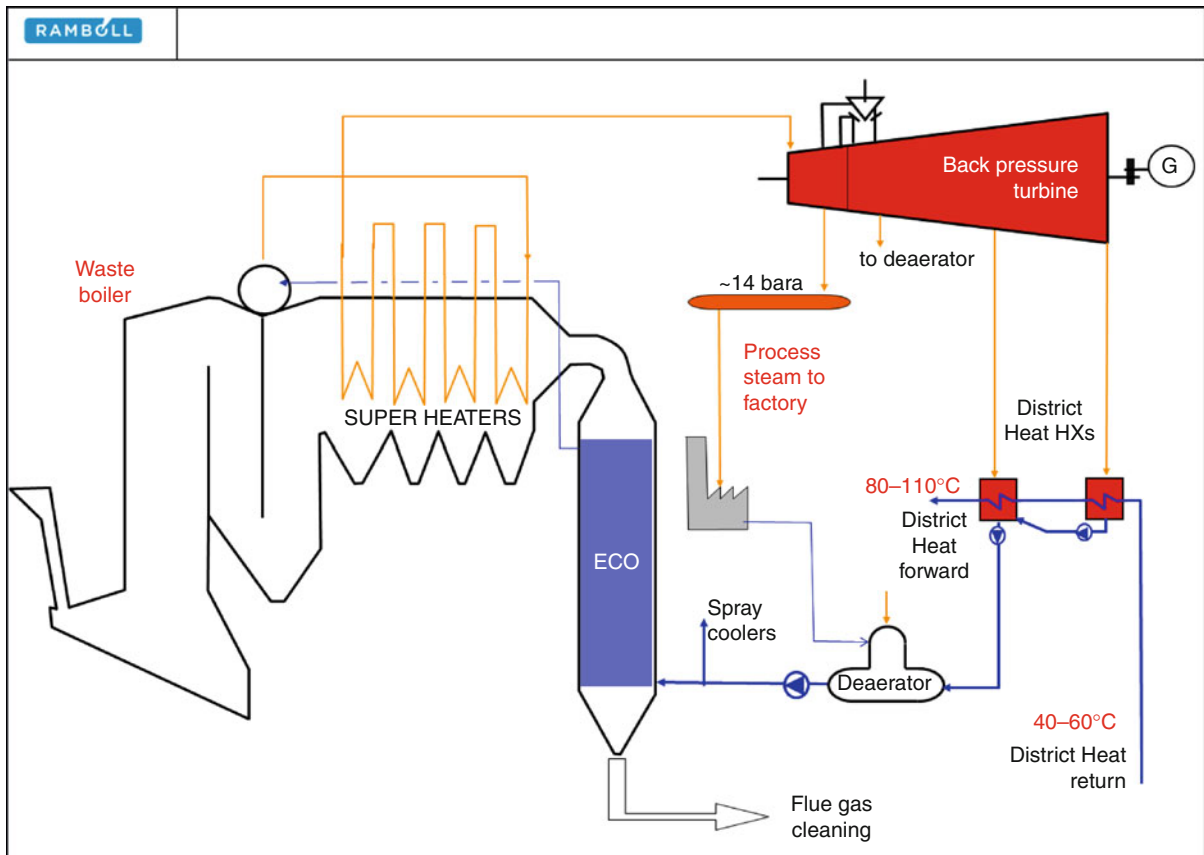
Waste-to-Energy for District Heating. Figure 13

1-2-3 rule of thumb for “district heat” configured waste-to-energy facilities. 1 t of waste yields approx. 2 MWh of heat, and 2/3 MWh of electricity

The district heat pumps would respond by decreasing the flow, to again obtain the desired district heat supply temperature. In such a system, the heat network is relying on other production units to balance the heat demand and heat production at any given point in time.

It is common, during summer, that a waste back pressure turbine produces more heat than the heat network demands. Since the MSW must be processed in the WTE plant, it is not always possible to reduce the thermal boiler load. Instead, plants often install water-to-air “summer coolers” to cool away the excess district heat.

Ultimately, the size of the district heating network, as compared to the size of the waste-to-energy facility,



Waste-to-Energy for District Heating. Figure 14

Back pressure turbine for district heat production. Such a setup does not require a (large requiring) air-cooled condenser, since all the heat is transferred to the district heating network

Waste-to-Energy for District Heating. Table 2 Estimates of electrical and total efficiencies for “district heating” versus “electricity-only” Energy for waste facilities. Assumptions: electricity only back pressure = 100 mbar. District heating network temperatures = 50–90°C. Live steam parameters: 425°C/52 bar

Size: 100 MW thermal 36 t/h of waste (~300,000 t/year) CV = 10 MJ/kg	Net electrical efficiency (example)	Heat efficiency (example)	Total efficiency (example)
Back pressure (district heat) plant (without air-cooled condenser)	22.5% (22.5 MW sold)	63% (63 MJ/s heat sold)	85.5%
Electricity-only plant (with air-cooled condenser)	26% (26 MW sold)	0%	26%

will determine whether it is more cost-effective to install an extraction turbine with an air-cooled condenser, or a district heating back pressure turbine with a water-to-air hot water cooler.

Advanced Solutions: Heat Pumps and District Cooling

Waste-to-energy plants (and other thermal installations) can be equipped to supply both district or industrial heating and cooling. This could be a preferred solution for a geographic location with cold winters but hot summers. For this application, an electrically or steam-driven chiller could be used. The electricity or driving steam could be supplied by the waste-to-energy steam turbine, thus allowing for the complete installation to supply power, district heat, and district cooling, depending on the current demands.

A schematic diagram of such “tri-generation” is shown in Fig. 17, where the “generator package” is driven by a waste-fired steam turbine. District cooling is less common than district heating, and in most cases only makes sense in hot countries, or for supply of industrial cooling to large office buildings, airports, etc. District cooling on household level is not common.

Optimizing Energy for Waste Facilities for CHP Operation

A WTE facility typically has efficiencies in the following order of magnitude (MWt: thermal input) (Ramboll calculations and experience from Ramboll projects. Some plants may deviate from these values):

“Electricity-only” WTE facility up to 100 MWt	23–26% net electrical efficiency
“Electricity-only” WTE facility > 100 MWt	Up to 30% net electrical efficiency
Heat and power WTE facility up to 100 MWt	17–23% net electrical efficiency
Heat and power WTE facility > 100 MWt	Up to 25% net electrical efficiency

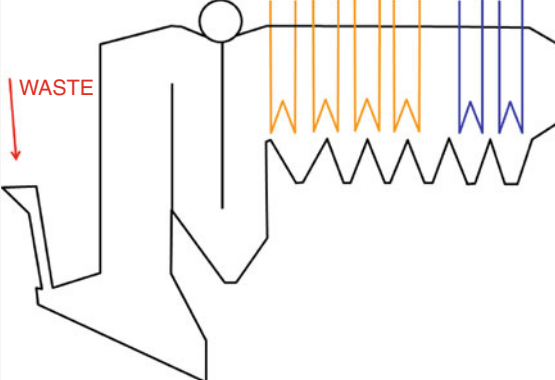
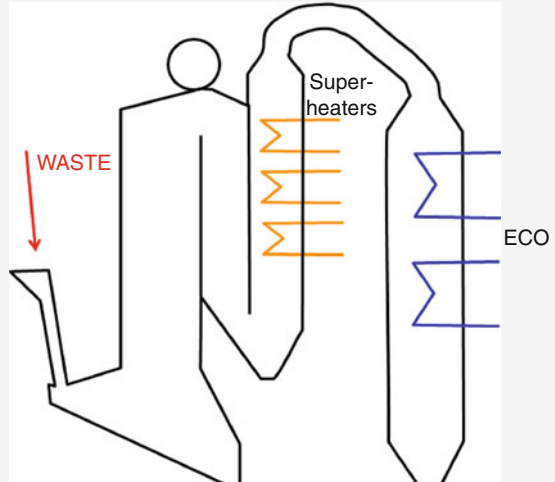
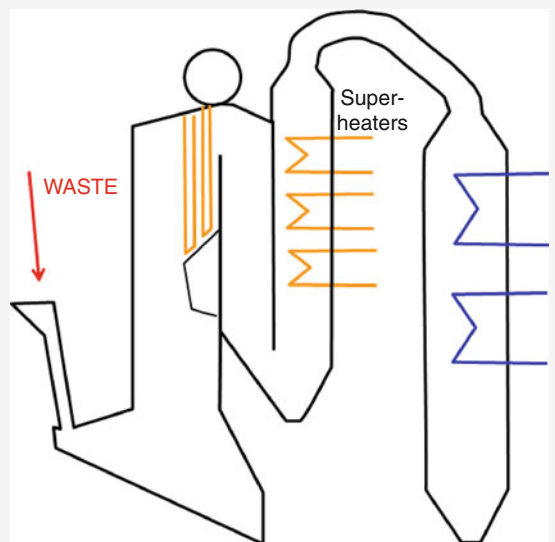
The lower efficiency of combined heat and power facilities is due to a higher turbine back pressure, as discussed earlier (Fig. 6). However, combined heat and power facilities have total efficiencies ranging from 82% to 86%; hence, a large amount of the energy in the waste is recovered for useful purposes.

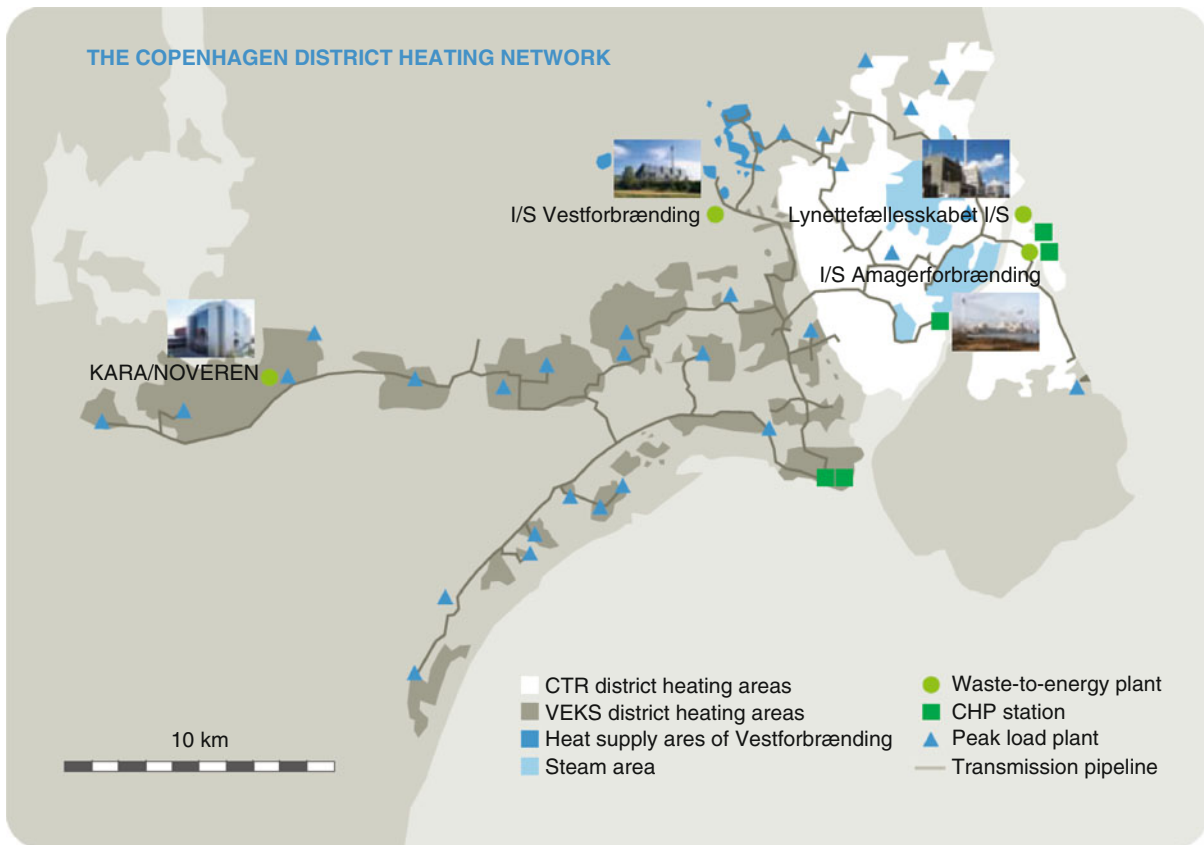
WTE facilities can also be equipped with so-called flue gas condensation equipment, where the boiler flue gas is cooled to 30–50°C to produce additional district heat. This flue gas temperature is below the flue gas dew point of normally 52–60°C, and hence, moisture in the flue gas is condensed, causing a significant energy release (The flue gas dew point depends on excess air ratio and absolute pressure, as well as the moisture content in the flue gas. The latter is closely related to waste lower heating value (LHV), since a lower heating value usually arises from a wetter fuel. A dew point of 51°C corresponds roughly to an LHV of 12 GJ/t, and a dew point 60°C corresponds roughly to an LHV of 8 GJ/t).

Facilities using flue gas condensation may reach total energy recovery efficiencies of >100% (measured on lower heating value basis), and are thus extremely efficient.

In relation to optimizing electricity output for a combined heat and power plant, the temperature of

Waste-to-Energy for District Heating. Table 3 Typical boiler Configurations

	<i>"Horizontal" waste boiler</i> Often used in Europe for larger WTE facilities Advantages: – High efficiency, high availability – Good access to cleaning tubes in fourth pass, hence lower maintenance costs Disadvantages: – Higher initial investment
	<i>"Vertical" waste boiler</i> Often used in smaller WTE facilities Advantages: – Low initial investment (smaller boiler) – Smaller plant footprint Disadvantages: – Replacement of superheaters, and in general maintenance, more difficult due to poorer access conditions – Economizer tubes cleaning by soot blower (= > lower efficiency)
	<i>"Vertical" waste boiler with superheater tubes in the radiation pass</i> This boiler configuration is often seen in USA Advantages: – As the "vertical" boiler, although even more compact Disadvantages: – The superheater tubes in the radiation pass may incur additional risk of corrosion and hence maintenance costs



Waste-to-Energy for District Heating. Figure 15

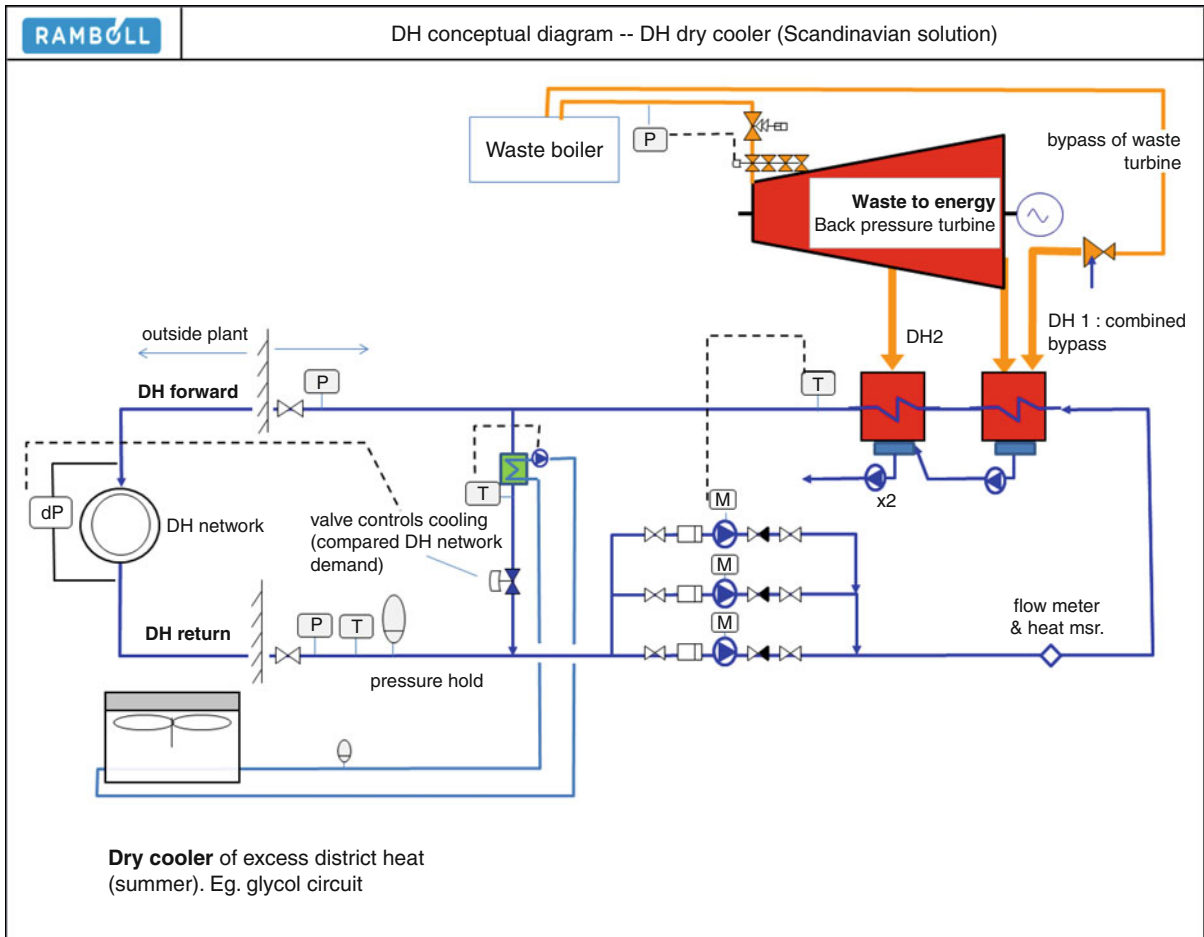
Copenhagen district heating network spanning about 40 km from east to west (Figure courtesy of Ramboll Waste-to-Energy)

the district heating water (especially the supply temperature) has a big influence on efficiency, as can be seen in Fig. 18. The figure shows that the electrical efficiency is reduced by approximately 1%-point for every 10°C increase in hot water supply temperature, which compared to other measures is a strong impact. A lower district heating supply temperature also results in a lower heat loss in the network, so there are many benefits to be gained by striving toward the lowest possible supply temperature. On the other hand, as discussed earlier, a higher supply temperature will allow transfer of more heat (or a reduced water flowrate) in the district heating pipework, which is also bound by hydraulic limits. It is recommended by Ramboll that for new district heating networks, the design temperatures should be a focus point, among others, when optimizing the overall system. For

existing districting networks, it is difficult to change supply and return temperatures, due to the coupling to existing pipe sizes, hydraulics, and final user installation.

Even for relatively big changes in district heating temperatures, which would result in large/expensive impacts on the mechanical design, the effect on electrical efficiency is only within 1–5% points. Other optimization measures toward higher power output, such as higher steam parameters, higher boiler efficiency, etc., influence electrical efficiency also in the range of 1–5% points, at the most.

These relatively slight changes in electrical efficiency are in stark contrast to the benefits of combined heat and power, where the plant total efficiency can be increased by up to 60% points (from 25% to 85%) by using the “waste” heat for useful purposes, if possible.



Waste-to-Energy for District Heating. Figure 16

Example of a district heat conceptual diagram for a “heat-only” back pressure plant

Obviously, heat and power requires a heat sales opportunity, which in practice means an energy infrastructure of district heating pipework and pump installations. The district heat network is not cost-effective on the short term, and certainly not for the individual producers of heat. However, municipalities, large energy companies, or governments interested in the long-term benefits of highly efficient energy facilities (for all fuels) can drive forward such infrastructure investments to the benefit of the whole community.

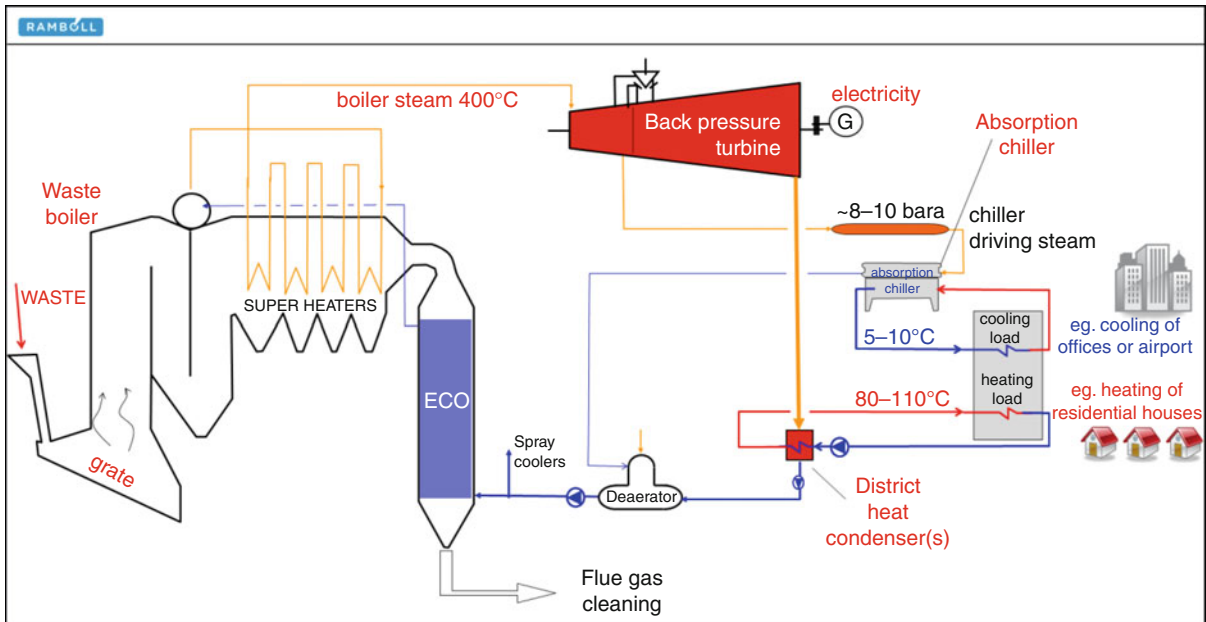
When using flue gas condensation in combination with a standard heat and power-configured plant, the total increase in efficiency can be up to 80–85%-points, resulting in total energy recovery rates above 100% (based on lower heating value).

Figure 19 illustrates the potential for flue gas condensation: up to approximately 20 MJ/s additional heat output for a 90MWt plant. As can be seen, the output is very sensitive toward the district heating return water, since the water is used as the media to cool the flue gas below the dew point of 52–60°C.

It is possible, and relatively common in Scandinavia, to incorporate absorption chillers to thus boost output of the flue gas condensation installation, if the district heat return temperature is high.

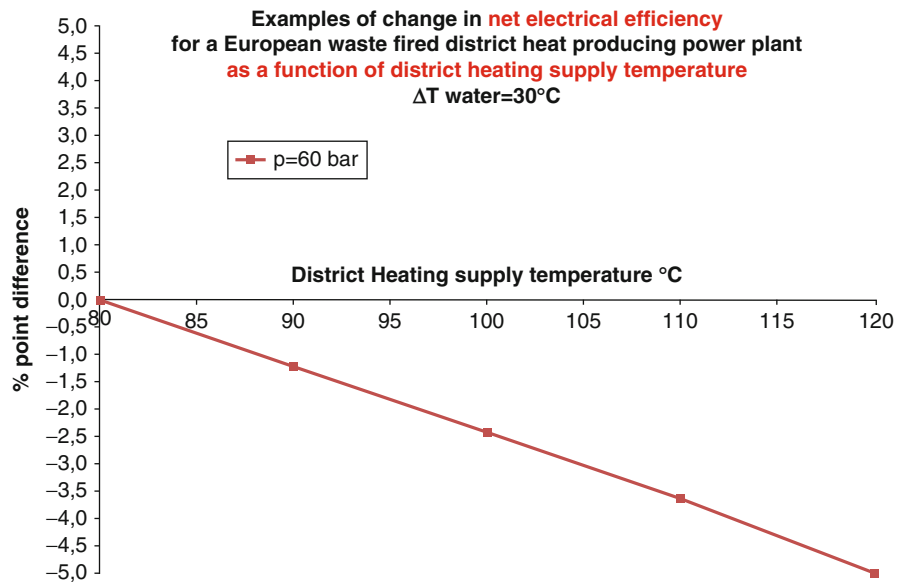
Conclusions and Future Directions

If a heat market exists or can be made available, waste-to-energy facilities can be a cheap source of



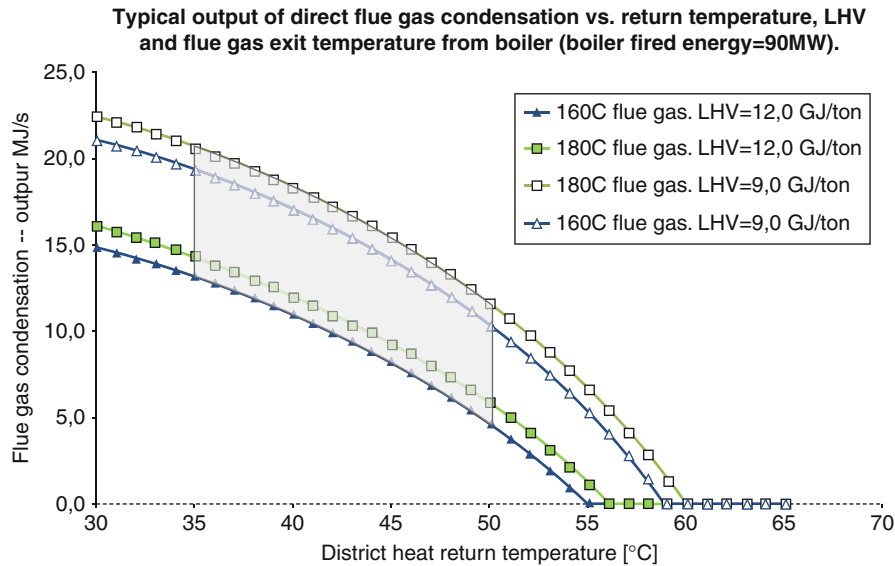
Waste-to-Energy for District Heating. Figure 17

Generation of power, heat, and cooling: "tri-generation". Cooling provided by an absorption chiller



Waste-to-Energy for District Heating. Figure 18

Change in net electrical efficiency for a 100 MW WTE facility as a function of the supply temperature in the district heating network (Source: Ramboll calculations)



Waste-to-Energy for District Heating. Figure 19

Additional heat output from flue gas condensation for a 90 MJ/s plant. The additional increase in total efficiency is in the range of 0–20% (Source: Ramboll calculations)

heat – for example, the total energy recovery for “electricity-only” facilities of about 25% can be increased to 85–105% for combined heat and power installations. The loss in electrical efficiency by doing so is moderate: The electrical efficiency is reduced by only 3–6%-points.

Furthermore, configuring a steam turbine for “heat and power” instead of “heat-only” generates heat with high efficiency: an “electricity-to-heat” ratio of up to almost 20!

Even for smaller amounts of heat extraction from a waste plant, it is not uncommon to achieve an “electricity-to-heat” ratio of up to 10. Using such heat is particularly interesting given the fact that waste incineration continuously takes place mainly to reduce the mass and volume of (nonrecyclable) waste, and, hence, the alternative is to reject large amounts of heat in cooling installations with no real benefit.

Bibliography

Primary Literature

1. Kleis H, Dalager S (2004) 100 years of waste incineration in Denmark. Babcock and Wilcox, Vølund and Ramboll, Copenhagen

2. <http://www.ens.dk/Documents/Faktaark/affald%20150709.pdf>
3. Phyllis database www.ecn.nl/phyllis and “Accomplishments from IEA Bioenergy Task 23: Energy from Thermal Conversion of MSW and RDF” 2001 and “21st Century Advanced Concept for Waste-Fired Power Plants.” Babcock Wilcox Volund
4. 21st Century advanced concept for waste-fired power plants. Babcock Wilcox Volund, p 20

Books and Reviews

- Afval Energi Bedrijf, City of Amsterdam – Waste and Energy Company (2006) Value from waste, waste fired power plant. The new standard for recovery of sustainable energy, metals and building materials from urban waste
- IEA Bioenergy (2000) Accomplishments from IEA Bioenergy Task 32: energy from thermal conversion of MSW and RDF
- Danish Board of District Heating (2006) News from DBDH, Energy Environment. Journal number 4/2006. Theme issue: waste and district heating
- Hesseling WFM, Rademakers PLF (2003) TNO environment, energy and process innovation, March 2003 (R2003/127): efficiency increase of waste-to-energy plants, evaluation of experience with boiler corrosion and corrosion reduction
- Rand T, Haukoil J, Marxen U (2000) Municipal solid waste incineration – a decision maker’s guide. World Bank, Washington, DC
- Renosam and Ramboll (2006) The most efficient waste management system in Europe * Waste-to-energy in Denmark

Waste-to-Energy using Refuse-Derived Fuel

FLOYD HASSELRIIS¹, PATRICK F. MAHONEY²

¹Hasselriis Associates, Forest Hills, NY, USA

²Energy Answers International, Inc., Albany, NY, USA

Article Outline

Glossary

Definition of Subject

Introduction

Mass-Burn Incineration: Predominant Method for Waste Combustion

Early Efforts to Recover Materials and Energy from Processed Refuse

Advanced Resource Recovery Facilities

Processing Equipment

Air Quality Control Equipment

Properties of Municipal Waste and Processed Refuse Fuel

Power Production from Refuse Fuel

Future Directions

Bibliography

Glossary

Mass-burn incineration Mass-burn incineration is a waste treatment process used in many locations throughout the world that involves the combustion of unsorted, unprocessed mixed waste on a moving grate, generally with recovery of the heat generated.

Municipal solid waste (MSW) MSW is all the garbage, including organics like grass and leaves, generated by households, commercial sites (restaurants, stores, offices, etc.), and institutions (schools, museums, public parks, etc.). Materials such as packaging and office paper from small- to mid-size factories, called “light industrial sites,” are also part of this mix.

Processed refuse fuel (PRF) PRF is a fuel created by the resource recovery system developed by Energy Answers International with minimum processing.

Refuse-derived fuel (RDF) RDF is produced from MSW through a number of processes to meet

requirements for particle size, moisture content and noncombustible content dictated by the thermal unit that will receive the RDF. At its simplest, MSW is shredded to a maximum particle size to produce RDF. More often, additional steps are taken to remove noncombustible materials and control the particle size.

Renewable energy Renewable energy is energy generated from fuel sources that can be replenished. Unlike fossil fuels, which are exhaustible, renewable energy sources regenerate and can be sustained indefinitely. The five renewable sources used most often are: Biomass – including wood and wood waste, municipal solid waste, landfill gas, and biogas, ethanol, and biodiesel; water (hydropower); geothermal; wind and solar.

Sustainable energy Sustainable energy is energy that meets the needs of the present without compromising the ability of future generations to meet their needs. Sustainable energy sources are most often regarded as including all renewable sources.

Definition of Subject

Combustion of municipal refuse to produce energy and recover recyclable materials is consistent with goals to diversify energy generation sources and implement sustainable and renewable options. Using waste as a fuel can decrease dependency on non-sustainable fossil fuels, such as coal, oil, and gas, and provide a consistent and reliable source of recyclable materials, as well as get rid of the waste and minimize landfilling.

Processing the refuse to generate energy and recover materials before and after burning is an approach that has been provided as an alternative to mass-burn incineration. This alternative has been practiced as a parallel choice since the early 1970s, and, in many situations, has provided advantages in the form of increased energy production, lower landfill disposal requirements, higher quality combustion residues with low unburned carbon content, complete combustion, cleaner gases going to the air quality control system, higher materials recovery rates, lower capital costs, and lower operating costs.

With growing concerns related to greenhouse gas production, environmental impacts of landfills, and

dependency on foreign oil, as well as a shift toward renewable and diversified energy sources, processed refuse fuel technologies should see an increase in demand.

Introduction

Starting in the mid-1800s incinerators were used for the primary purpose of reducing waste volume. These incinerators, which burned the waste “as is” with no preprocessing, were widely used and were successful at reducing waste transport and disposal costs for densely populated urban areas. Numerous grate systems were designed and installed to burn the waste and move the ash residue out of the incinerator for disposal.

Early efforts to utilize the heat produced as motive force to operate in-plant equipment and to generate low-pressure steam for district heating or industry was the next step which began in many European locations. Urban areas in the USA, where land was becoming more valuable, followed suit as documented in the August 28, 1897 issue of the *Scientific America Magazine* featuring a heat recovery incinerator in operation in New York City producing electric power used for lighting the Williamsburg Bridge. However, despite waste volume reduction of up to 80%, energy recovery was an insignificant and inefficient part of the incineration solution and, because incinerators were operated with virtually no air pollution control systems, they were a major contributor of poor air quality.

Mass-burn incineration technology advanced in the mid-1900s with the addition of more efficient heat recovery units and, in some cases, equipment to recover a portion of the metals in the waste.

An important milestone in waste management strategy for the USA was the passage of the Resource Conservation and Recovery Act (RCRA) in 1976. As stated in Section 1003, “The objectives of the Act are to promote the protection of health and the environment and to conserve valuable material and energy resources. . .” The National Policy is presented in Section 4001 as “. . . developing and encouraging methods for the disposal of solid waste which are environmentally sound and which maximize the utilization of valuable resources including energy and materials which are recoverable from solid waste and to encourage resource conservation.”

Another important milestone was the development of the Energy Policy and Conservation Act of 1975. This Act called for actions to encourage the expansion of domestic energy sources. Conversion of municipal solid waste (MSW) to energy is a noted alternative.

These government policies prompted regulatory agencies to demand closure of most of the old incinerators operating in the USA and required the addition of air pollution control systems on newer units. These policies also provided a strong foundation for advancing the technologies to process refuse for improved recovery of materials and energy.

With the world experiencing its first “energy crisis” in 1973, marked by the tripling of petroleum prices, a new opportunity for waste management engineers emerged. Waste, refuse, garbage, trash, etc., with a 75% combustible fraction, represented a potential source of energy which engineers could look to recover and reduce the cost of disposal. This historical event began the change in the design objectives of waste management engineers from waste reduction with some energy recovery to maximum resource recovery and efficient energy generation. Waste was now starting to be viewed as a “resource” and alternative fuel source.

Mass-Burn Incineration: Predominant Method for Waste Combustion

Mass-Burn incineration systems are the predominant method for combusting waste to produce energy. Combustion systems designed to process waste into a fuel in order to maximize energy generation and materials recovery have a distinctively different design philosophy and use different equipment. A brief description of mass-burn incineration and some of the disadvantages of the technology provides a foundation and rationale for the processing of refuse to recover materials and energy from refuse.

At a mass-burn incinerator, Municipal Solid Waste (MSW) is discharged from transfer vehicles or waste collection vehicles into a deep pit. There is no processing of the waste. Waste is removed from the pit by overhead crane and fed into hoppers which feed the furnace. Most of the time, the crane is used to mix the waste to try to obtain uniformity

of heating value, although this is somewhat difficult to do due to the distance from the crane cab to the waste itself.

In order to ensure continuous operation, it is necessary to have two cranes, while only one at a time is used. Crane cable replacement is a continuing and significant source of maintenance cost (Fig. 1).

The raw waste falls onto a traveling grate which is designed not only to move the waste from front to back (usually a distance of some 60 ft, divided into three 20-foot-long grates with drops from one to the next), but also to tumble the material in order to provide oxygen necessary for combustion. The grate size is determined by the amount of heat release anticipated, commonly about 200,000 BTU per square foot of grate per hour.

In order to facilitate movement of MSW through the furnace, the grate slopes downward from front to back such that the overall height of the grate is about 20 ft. This adds 20 ft to the overall height of the boiler house.

Because there is no presorting of material being fed to the furnace, there is always the problem of pieces of metal, cables, water tanks, propane tanks, etc., being fed into the furnace. This creates a potential erosion problem and therefore the waterwalls of the furnace are covered with refractory linings to protect the tubes. This refractory lining is carried to a point very high in the furnace to avoid corrosion at the point where the refractory stops and bare tubes are exposed. The use of refractory causes slower and reduced heat transfer into the tubes while slag from melting glass and low melting point metals tends to build up on the refractory linings low in the furnace. Thus, on a regular basis, the boiler must be shut down and refractory either cleaned or replaced.

As all the waste is burned on the grate, temperatures on the grate are above the melting point of glass such that clinkers are formed. On discharge from the furnace, the ash falls into a trough full of water where the clinkers shatter into small particles. These particles are amalgams of the noncombustibles originally fed into the furnace along with unburned material (carbon particles). Appearing often in the ash from mass-burning units may be unburned items such as large telephone directories, layered floor polisher pads, watermelon and grapefruit rinds, and other dense organics and similar materials which do not burn well because the

necessary combustion air cannot reach in between the layers of paper or cloth. Because large metal objects are in the ash, plugging of the ash discharge systems can occur. Although the removal of some ferrous metal from the ash is often easily accomplished with magnets, mechanical removal and recycling of other metals and valuable products is virtually impossible. The use of water quenching for the ash produces an alkaline solution which generally must then be treated and neutralized.

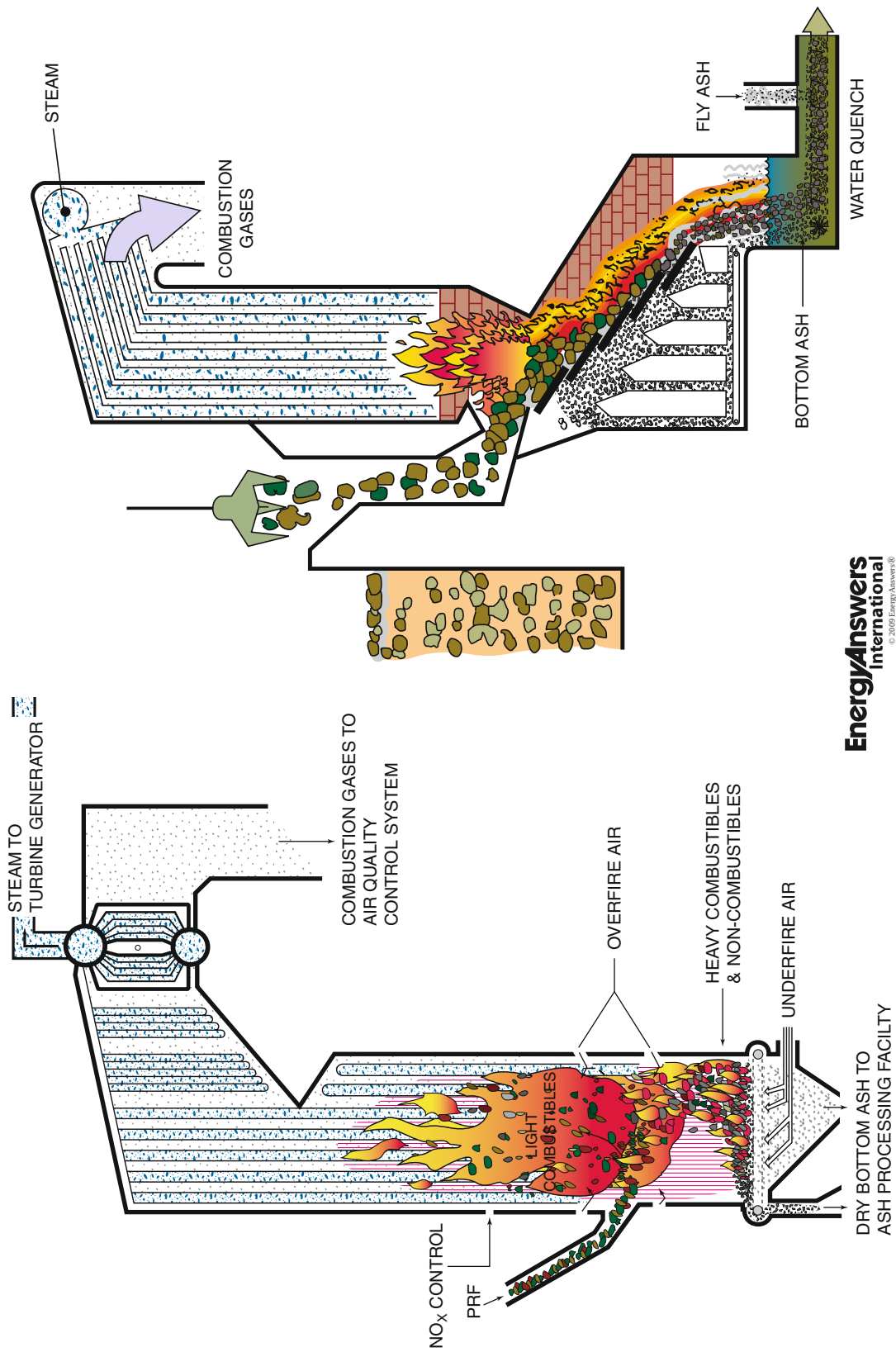
Early Efforts to Recover Materials and Energy from Processed Refuse

Although mass-burn incinerators are viable methods for generating energy from waste, they are inherently inefficient for recovery of recyclable material. In the early 1970s technology emerged for processing waste to increase the recovery rates for energy and materials. Many of the early designs were flawed and were shut down but provide an understanding of the history and evolution of the technology which is currently implemented at the highly successful projects in operation or in development in 2010.

National Center for Resource Recovery (NCRR)

In the early 1970s, one of the concepts promoted by the newly formed “environmental” movement in the United States was that beverage containers made from metals and glass should all be reusable, thus reducing the amounts of new glass and metal which would have to be produced to meet the needs of future consumers. Container manufacturers, however, had only recently discontinued the use of “refillables” and were concerned that the costs of collecting, separating, returning, sterilizing, and reusing of containers would be very high compared to the cost of new containers. The National Center for Resource Recovery (NCRR) was formed and financed by several container manufacturers, aiming to demonstrate that recyclable material could be recovered from MSW at a central processing station and that this material could then be used to manufacture new containers and would negate the need for returnable containers.

Complications arose due to the fact that MSW contains not only containers but a wide variety of other waste materials including, but not necessarily



Waste-to-Energy using Refuse-Derived Fuel. Figure 1
Processed Refuse Fuel Combustion compared with Mass Burn Combustion. [Energy Answers International]

limited to, food waste, paper and cardboard, magazines, household chemicals, old clothes and rags, various plastic products (including containers), garden vegetation, trees and brush, flammable materials, sand and dirt, etc. What developed at NCRR was a series of systems and equipment aimed at separating not only containers, but also any other products which might be useful, including materials which might be used as fuel in place of coal or other fossil fuels, noncombustible materials (sand, grit, etc.), organic matter to create a soil conditioner; and valuable materials (metals, coins, jewelry, and the like).

One of the products which the NCRR tried to create was a “Refuse-Derived Fuel” (RDF) which could be used in existing furnaces and boilers as a substitute for, or in combination with, coal, oil, or gas [1]. The first stage was to separate the light combustible material (paper and light plastics) from the non-combustible materials by chopping all of the material into smaller pieces in a shredder and then separating the light fraction from the heavier material by means of air currents. Unfortunately, less than half the material became “air-classified” light combustible material while the balance, over 50% of the incoming waste, was destined to be disposed of in landfill. A further problem that existed in the preparation of this “fluff” RDF was the inability of equipment to handle the rags and streamers that occurred in shredded MSW. Various systems for the separation of these materials were tried with little long-term success.

Air-classified RDF did not prove to be a satisfactory fuel for use in facilities designed for fossil fuels. It was difficult to feed into the furnace, could cause fouling of flue gas paths, and, because of the high ground-glass content, significant slagging occurred on the grates and on the furnace walls. And because of the low density of the material, it was difficult and expensive to store and transport it.

The idea of creating very dense “pellets” made from MSW was born to solve the problems inherent in fluff RDF. MSW was ground into extremely small particles and a binder was added to enable the material to be extruded into pellets. Although this solved the high costs of storage and transport, the cost of manufacturing the pellets could not compete with the cost of normal fossil fuels. A further disadvantage was that the material remaining after production of pellets, at

least 40% of the original incoming waste, was sent to a landfill.

Virtually all the systems developed at NCRR ultimately proved to be either technically impractical or economically unfeasible due to lack of markets capable of buying the recovered materials at a price necessary to cover NCRR’s costs. Despite the unsatisfactory economics of the processes developed at NCRR, the facility operated for several years until finally the backers of NCRR withdrew financing and the facility has been closed since the 1980s.

Solid Waste Reduction Unit (SWARU)

In 1968, the City of Hamilton, Ontario, Canada, retained Gordon L. Sutin and Associates Limited, Consulting Engineers, and Consoer Townsend Engineers to design an incinerator which would burn 600 t/day of MSW, including a portion of combustible demolition waste [2].

SWARU was designed with a radically different philosophy in comparison to the combustion facilities in operation. Instead of designing a facility for burning waste, it was determined that the facility would become a manufacturing plant in which the raw material, municipal solid waste and demolition material, would be processed to create a finished product, ash. It was recognized that there would be by-products in the form of heat and air emission and possibly contaminated water. When this hypothesis was considered, it became clear that the problem was not really a “combustion” problem but rather a problem of receiving a “difficult-to-handle” raw material, feeding into a process for recovery of the heat generated in the combustion process, conversion to ash, removal of the finished product from the process, and control of air and water contamination.

An obvious result was the recognition of the need to handle the raw material and the finished product by the most simple and well-proven method known – conveyors. However, because of the nature of the raw material, movement on a conveyor reliably and without excessive belt damage could not be accomplished unless the waste could be converted to a more homogeneous mixture.

A literature search at the time indicated that only one large-scale, modern MSW combustion plant had

been built in Canada or the United States. The facility at the Norfolk Navy Base burned waste to generate steam but did not have any other recovery enhancement. However, many modern facilities had been built in Western Europe during the late 1950s and the 1960s that not only burned waste but also generated useful energy in the form of steam and/or electricity.

An evaluation of installations and equipment in Western Europe confirmed that none of the facilities processed MSW before burning. All technologies in Europe at that time were based on "Mass-Burn" techniques where, apart from the removal of large objects, all the waste was burned as received. There were significant components of the design which caused high cost and operational inefficiencies:

1. The use of the crane as a method of delivering waste to the furnace was expensive and inefficient with backup operators and crane equipment on standby to ensure continuous operation 24 h/day, 7 days/week.
2. The crane was also used, ineffectively, to reposition waste in the pit and to mix wet and dry waste in order to maintain a uniform heating value.
3. Despite the advantages of allowing for a large amount of waste to be stored in a small area with simplified odor control, the pit, which was in continuous use, could rarely be cleaned.
4. Different crane loads of waste had very different heating value and moisture content. Thus portions of the waste would burn more quickly than others and the result would be that where the waste had burned quickly, air flow through the grate to improve combustion and to cool the grate would be concentrated on those areas with the least resistance (where the waste had burned off quickly) while in other areas, there would not be enough air to support combustion. In these starved air areas, the waste would "cook" instead of burning, resulting in large clinkers and large percentages of unburned combustible materials.
5. Dense objects such as telephone directories and layered floor polisher pads could not be completely burned because char on the outside would act as an insulator and would block the flow of air necessary to burn all the materials. Thus, the job of burning all combustible material was not completely

achieved, resulting in high combustible content in the ashes.

6. The temperature of the refractory linings would ultimately reach the melting point of glass, and glass in the waste would melt onto the linings, drip to the cooler bottom of the furnace, and harden on the refractory to form slag that needed to be removed on a regular basis using jackhammers and other heavy equipment.
7. Owing to the nature of the waste, a common occurrence was the breaking of a grate bar due to jamming of small metal pieces between adjacent bars. Such a failure would result in burning material falling into the air hoppers beneath the grate and causing damage.
8. Because many large metal objects (hot water tanks, bicycles, etc.) would often be fed to the furnaces by the crane, it was not uncommon for these objects to get caught at the discharge end of the grate just above the ash hopper. Thus, the ash would be restricted from entering the hopper and major plug-ups could occur, requiring the unit to be shut down while the plug-up was cleared.
9. The presence of large quantities of metals and other non-friable materials inevitably caused high wear on the drag chains used to remove the ash from the water-filled ash trough. Stretching and ultimate breaking of chain links caused jams of the equipment, necessitating repairs on an emergency basis. Further, it was noted that the wet ash systems used large quantities of water, and maintenance of workable conditions for employees in the ash receiving room was very difficult due to high moisture conditions.

As a result of the observations made of European equipment, it was determined that the design of the SWARU would not be based on the mass-burn technology but rather would be a design which would eliminate, if possible, all the problems common to the mass-burn facilities. Thus, the aim was to avoid the following systems and methods:

1. The use of cranes for moving waste from storage into the combustion system
2. The use of highly complex mechanical grate systems

3. The use of refractory linings on the walls of the furnace
4. Water quenching of the ash

SWARU was the first successful waste combustion system that recovered materials and energy from processed refuse. The operation was purposely designed to be simple and avoid the problems associated with more complex RDF systems.

- Instead of receiving and storing the incoming waste in a pit, the material was dumped on the floor of an enclosed building, examined so that undesirable objects could be removed before processing, and fed onto conveyors leading to the shredders.
- The actual processing of the waste into fuel consisted only of manual removal of objects from the waste which might jam a shredder or conveyors, and then shredding the waste, creating a relatively uniform, homogenous “Processed Refuse Fuel.”
- The PRF is fed by conveyor into chutes that open into the lower part of a Babcock and Wilcox spreader-stoker boiler, just above the grate [3]. At the bottom of the chutes are deflectors that divert the material upward and additional throw is provided by a blast of air. The light materials, paper, plastic, textile, etc., catch fire in midair while the heavier material, glass, metal and chunks of wood and other combustibles are thrown to the back of the furnace where they land on the grate. The grate moves forward very slowly while air is blown through the grate to maintain temperatures below the melting point of glass. It takes about 1 h for the grate to move from back to front and on the last third of the grate, little burning is taking place but the ash is being cooled by the under-grate air.
- As a result of the simple fuel preparation and the “suspension-fired” combustion system, the ash was shown to be almost completely free of unburned combustible material, less than 1½% compared to up to 8% found in the mass-burn ash of the time. Buildup of slag was eliminated on the furnace walls; no water was used for removal of ash; and the grate material did not have to be manufactured from special metal alloys because the “hot spots” on the grate commonly experienced in mass-burn systems were not created.

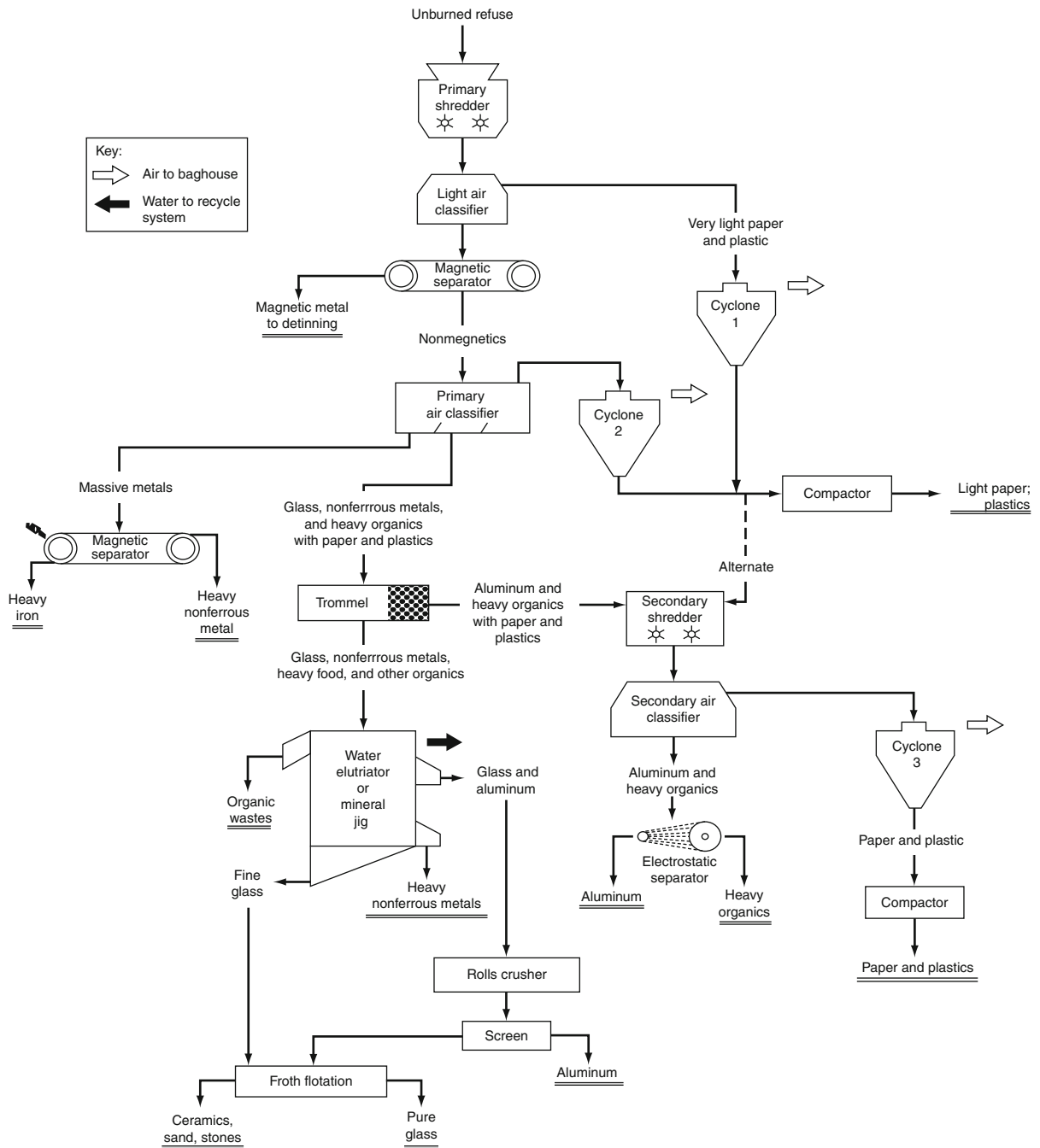
SWARU went online in 1972 and, after a period of operation, underwent substantial modification, then continued to operate for 30 years, when it was shut down due to changes in permissible emission limits and the costs associated with the required retrofits. This was the first “dedicated” RDF-firing plant, as distinguished from “co-firing” with coal [4, 5].

US Bureau of Mines

In the 1970s, the United States Bureau of Mines concentrated research on the recovery of materials in municipal waste – paper, cardboard, ferrous, aluminum, and glass. The pilot research plant at College Park, Maryland, included a shredder (flail mill), trommel screen, air classifier, and metals recovery equipment that recovered materials from samples of municipal waste obtained from various parts of the country (Fig. 2).

The next concept that arose was the use of RDF for co-firing in utility boilers. In 1974, St. Louis Electric took up this concept at their Merrimac Plant in Missouri [6, 7]. With funding from the US Environmental Protection Agency, a shredding plant was erected, with a system capable of feeding the RDF to the existing boilers. The EPA-sponsored research program analyzed the waste in detail, over the course of a year. The research program included sampling and analysis of the stack emissions while burning this fuel, co-fired with coal (Table 5). The RDF was able to supply 10–15% of the total energy input. After the trial runs, it was decided to install magnets after the shredding operation to recover ferrous material, which would cause problems in feeding of the fuel to the boilers. It was also decided to add a second stage of shredding to make the RDF more suitable for feeding to the boilers.

Burning this low-sulfur RDF with coal significantly reduced the sulfur dioxide emissions, as expected. However, emissions from co-combustion included higher levels of heavy metals, such as lead and cadmium, than with firing only coal. The existing boilers were permitted for the emissions with coal firing; new compliance tests would be required for co-combustion. This would abrogate the “grandfathered” existing permits, which the utility was reluctant to do. Plans to build a processing facility taking waste from



Waste-to-Energy using Refuse-Derived Fuel. Figure 2

US Bureau of Mines System (From [1])

the surrounding community were made and presented to the public. The public uproar convinced the utility that this would damage its public image.

In succeeding years, the US EPA sponsored extensive research on shredding, screening and air classification of RDF. Much of this work has been summarized in

Processing Refuse-Derived Fuel by Floyd Hasselriis, 1983 [1].

Co-Firing Operations

The City of Ames, Iowa, started up the second plant to process and burn its refuse in 1975 [8]. This plant employed two stages of shredding and an air classifier, and included a “silo” to store the RDF and feed it to the boiler, co-firing with coal. After a year of operation, it was decided to replace the air classifier with rotary screens made by Rader.

Combustion Engineering proposed burning RDF in their “corner-fired” boilers in Madison, Wisconsin, and in Chicago, Illinois [8]. Americology designed the process with two-stage shredding and air classification: it was started up in 1974. A shredder explosion damaged the plant, resulting in isolating the primary shredder in a concrete “cage.” The plant was shut down in 1979. The City of Chicago designed a similar processing plant, which was started up in 1977. In this case, the RDF had to be conveyed the long distance from the plant to the boilers by pneumatically blowing it through 20-in. diameter pipes to the boilers. It was found that the blowers did not have sufficient capacity to convey the RDF without accumulating glassy sediment in the pipes.

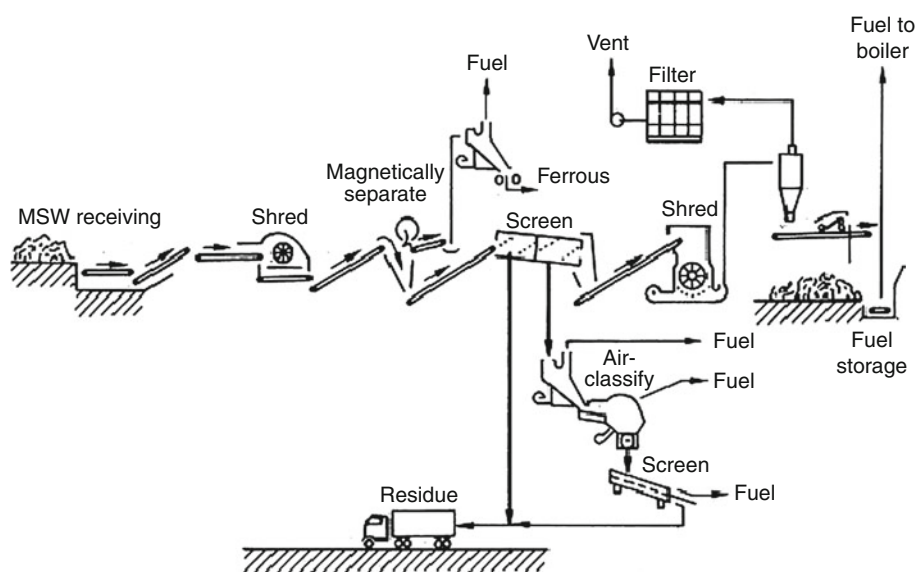
A series of plants were designed for co-firing of RDF with coal, in Akron, Ohio; Columbus, Ohio; Madison, Wisconsin; Monroe County, New York; and Lakeland, Florida, but were shut down for various economic, political, and technical reasons. The City of Columbus built a processing plant in 1982, to produce RDF to be burned in their municipal power plant. Co-firing the RDF with coal was not successful and high emission levels caused a shutdown of this facility.

Northern States Power built a waste processing plant in Red Wing, Minnesota, in 1987 that employed two stages of shredding and a Rader disk screen. The RDF production facilities in Wilmarth, Minnesota, and Lacrosse, Wisconsin, are operating today.

The difficulties of co-firing with coal contributed to the decision to build only “dedicated” boiler installations [9]. Combustion Engineering promoted plants in Hartford, Connecticut; Detroit, Michigan; and Honolulu, Hawaii, in 1988, all based on two-stage shredding, magnetic separation and trommeling (Fig. 3).

Albany NY Energy Recovery System (ANSWERS)

The Albany NY Energy Recovery System (ANSWERS), which went online in 1981, was modeled after the Processed Refuse Fuel system at the SWARU in



Waste-to-Energy using Refuse-Derived Fuel. Figure 3

Combustion engineering systems (From [8])

Hamilton, Ontario [2]. The ANSWERS process was divided into two sites. The City of Albany-owned site hosted a facility where waste was received and processed into Processed Refuse Fuel (PRF). At a second site, some 9 miles away, the PRF was burned to produce steam for the heating and cooling of State government office buildings in downtown Albany. The combustion facility was designed, owned, and operated by the State of New York Office of General Services, while the overall conceptual design of the entire system and the design and construction of the Processing Facility were by Smith and Mahoney, Engineers, of Albany, New York with Gordon Sutin acting as Technical Advisor and consultant to Smith and Mahoney. The steam plant burned PRF for 15 years until it was closed in order to avoid capital expenditures needed to meet new, more stringent air pollution control regulations.

As shown in Fig. 4, the municipal waste carters dumped their loads on the tipping floor. Front loaders moved the waste to the metal pan feed conveyors, where the waste was inspected and objectionable materials removed. The raw waste was then conveyed to the top of the shredders. Processing was done with two 1,000 HP diesel engine-driven shredders, in a single stage. After passing under magnets to remove ferrous material, the PRF was loaded into transfer trailers and shipped to the power plant where it was deposited into a pit for storage and feeding by crane to the boilers.

ANSWERS demonstrated the versatility of processed refuse fuel systems when waste can be processed into fuel at one location (or several locations if necessary) and transported by truck (or rail) to a power plant in a different location.

Bridgeport, Connecticut Eco-Fuel

In 1986, Combustion Equipment Associates built a plant to process MSW into a powder, called EcoFuel, to be transported and stored in silos, for co-firing with coal or oil in cyclone-furnace utility boilers at United Illuminating. The trash was shredded and ground in ball mills to a storable and retrievable powder. The process was demonstrated to be successful but could not compete with the low cost of landfills in the region. The project was shut down, as it was unable to contract with communities for the project's 2,300-t/day capacity [10].

The Southeastern Public Service Authority (SPSA)

The Southeastern Public Service Authority in Portsmouth, VA, built a facility in 1988 to produce RDF to be burned in the Naval Shipyard. This plant uses two-stage shredding, magnetic separation, and two stages of trommel screening to produce a fuel suitable for burning in the Navy boilers. Belt conveyors carry the fuel from the processing plant to the boilers where fuel feeders supply the fuel to the existing boilers (Table 1).

Advanced Resource Recovery Facilities

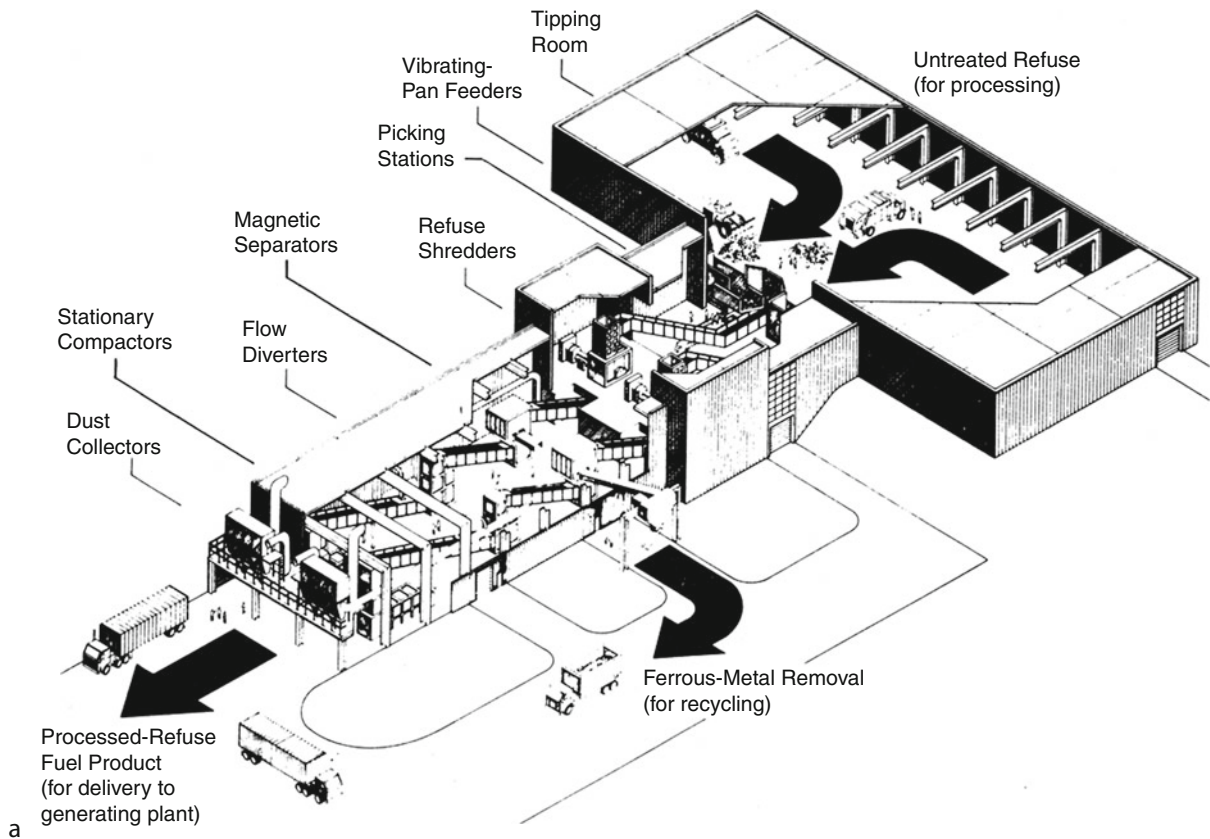
SEMASS Resource Recovery Facility

The SEMASS (Southeastern Massachusetts) Resource Recovery Facility in Rochester, MA, was developed by Energy Answers [2]. Construction began in 1986 and the facility went online in 1989, accepting 2,000 t/day of solid waste from communities on Cape Cod and in the southeastern section of Massachusetts. In 1996, SEMASS was expanded to 3,000 t/day with the addition of a third boiler, second turbine, and associated waste receiving, processing, and air quality control equipment.

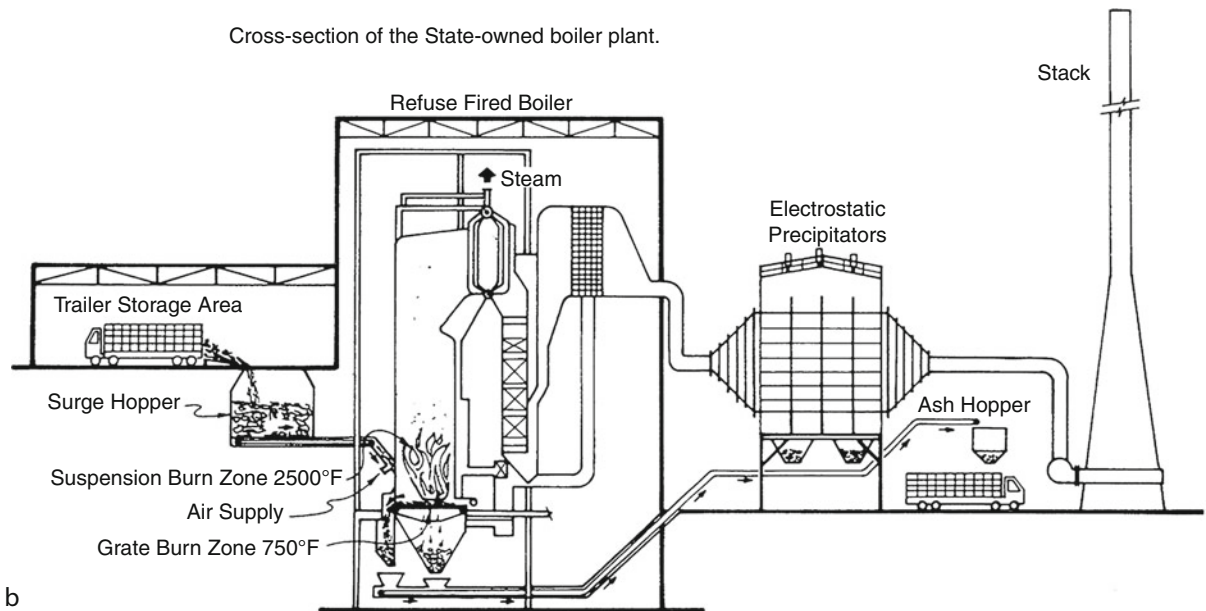
An aerial photo of the SEMASS Resource Recovery Facility plant is shown in Fig. 5.

The first step takes place at the municipal level where residents participate in programs to reduce the amount of waste by recycling and composting. Whatever waste cannot be diverted through those programs is delivered to SEMASS by packer trucks, transfer trailers, and rail cars (Figs. 6 and 7).

After crossing the scales, waste is received on a tipping floor, where an attendant watches for unacceptable materials and either refuses them or sets them aside. Unacceptable items are such things as large rolls of carpet or wire fencing, tires, tree trunks, and similar items that should not be processed in the shredders. The waste is scanned for radioactive content or combustible potential (fireworks, propane tanks). It is then delivered by loaders to conveyors feeding shredders that reduce the material to pieces of nominal maximum size of 15–20 cm. Most of the particles in the shredded waste are below 5 cm and friable materials such as glass are still smaller. A magnet removes ferrous material, which goes to metal scrap dealers. Although the Processed Refuse Fuel (PRF) is in fact



Cross-section of the State-owned boiler plant.

**Waste-to-Energy using Refuse-Derived Fuel. Figure 4**

(a) Refuse fuel processing facility, Albany New York; (b) Processed refuse-fuel combustion and power production facility, Albany, New York

Waste-to-Energy using Refuse-Derived Fuel. Table 1 Refuse processing power plants operating in the USA in 2007 [11, 12]

Plant name and location	Start date	Capacity (t/day)	Net power
Ames Municipal Electric Utility, Ames, IA	1975	1 line at 175 = 175 tpd	10 MW
Mid-Connecticut Resource Recovery Facility, Hartford CT	1987	3 lines at 666 = 2,000 tpd	68 MW
Miami-Dade County RRF, Miami FL	1979	4 lines at 750 = 3,000 tpd	77 MW
Maine Energy Recovery Company, Biddeford, ME	1987	2 lines at 300 = 600 tpd	22 MW
Penobscot Energy Recovery Company, Penobscott, ME	1988	2 lines at 750 = 1,500 tpd	25 MW
SEMASS RRF, Rochester, MA	1989	3 lines at 900 = 2,700 tpd	78 MW
Great River Energy – Elk River, MN	1989	2 lines at 250 tpd	35 MW
		1 at 500 = 1,000 tpd	
Xcel Energy – Red Wing Steam Plant, Red Wing, MN	1987	2 lines at 360 = 720 tpd	21 MW
Xcel Energy – Wilmarth Plant, Mankato, MN	1987	2 lines at 360 = 720 tpd	22 MW
Southeastern Public Service Authority, Portsmouth, VA	1988	4 lines at 500 = 2,000 tpd	50 MW
Xcel Energy French Island, LaCrosse, WI	1987	2 lines at 200 = 400 tpd	32 MW
North County Resource Recovery Facility – West Palm Beach, FL	1989	2 lines at 900 = 1,800 tpd	62 MW
Honolulu H-Power, Honolulu, HI	1990	2 lines at 925.5 = 1,851 tpd	57 MW
Greater Detroit Resource Recovery Facility, Detroit, IL	1991	3 lines at 944 tpd = 2,832	68 MW
Totals:		21,298 tpd	627 MW



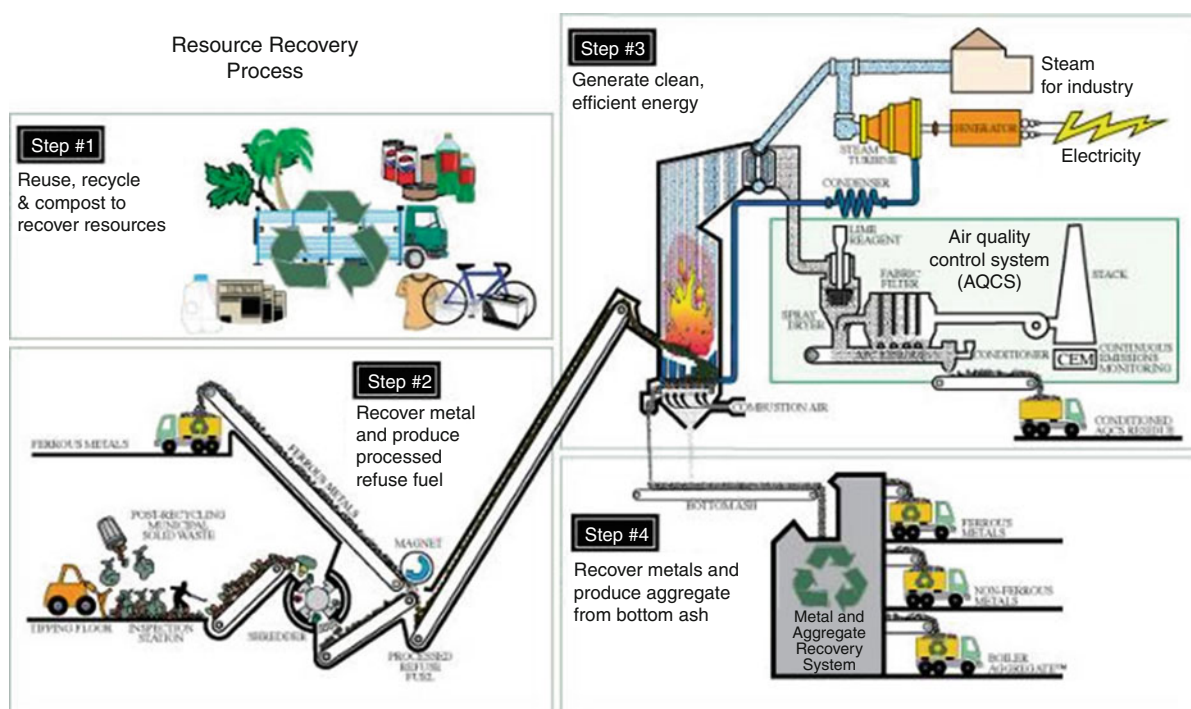
Waste-to-Energy using Refuse-Derived Fuel. Figure 5

Aerial view of Southeast Massachusetts resource recovery facility (SEMASS) (© 2010 Energy Answers International, Inc)



Waste-to-Energy using Refuse-Derived Fuel. Figure 6

Processed refuse fuel in SEMASS storage building; front loader moving load to conveyor ((© 2010 Energy Answers International, Inc)



Waste-to-Energy using Refuse-Derived Fuel. Figure 7

SEMASS schematic process diagram (© 2010 Energy Answers International, Inc)

a “Refuse-Derived Fuel,” the name PRF has been adopted to distinguish it from other Refuse-Derived Fuels such as “fluff” and “pellets,” which require extensive processing of the waste compared to the two-step processing developed by Energy Answers.

The PRF leaves the shredders and is conveyed to the boiler building. Here the boilers each draw off the

quantities needed to maintain steady steam generation. The PRF is fed into automatically controlled feed systems that blow the PRF into the furnace section of waterwall boilers. The steam produced drives the electric generators in the conventional power plant. Excess PRF is conveyed back to the storage floor, from which the loaders retrieve fuel as needed when the processing

floor is not operating, such as at night and on weekends.

Heavy materials drop onto a simple horizontal moving grate at the bottom of the furnace. The material on the grate consists of noncombustibles such as metal, glass, ceramic, and sand combined with small pieces of heavier combustible materials such as wood. The grate moves from back to front of the furnace over a 1-h period with a 20–30 cm deep bed on top. For the first 2/3 of the 6-m long grate, the heavy combustible materials burn on the grate. The grate temperatures are kept below the melting point of glass to avoid slag formation. Because of the long period allowed for the combustibles on the grate to burn, this “bottom ash” contains less than ½% carbon.

The lighter materials in the PRF go upward in the furnace and catch fire in midair forming a high-temperature “fireball” a few meters above the grate [3]. In this location, because of the ease with which the fuel can intermix with air, the amount of excess air which must be injected to complete combustion is a reduced and thus temperatures up to 1,250°C are reached. The reduced amount of excess air results in maximum thermal efficiency and therefore more electricity is generated per ton of waste than would normally be the case in a mass-burn system.

Bottom ash and fly ash are handled separately, so the fly ash can be treated to minimize the potential for leaching of heavy metals such as cadmium and lead. The bottom ash, which is generally granular in form, is discharged from the boiler without quenching in water, and is transported to the on-site Bottom Ash Processing Building. In this building, three products are recovered: the ferrous metals not removed prior to combustion, the nonferrous metals, and the Boiler Aggregate™, a granular material which can be substituted for natural aggregates used for landfill daily cover and as an ingredient in asphalt and concrete. The nonferrous metals are sold to a dealer who separates the contents of the mix, and the ferrous metals are sold to scrap dealers.

Because the heavy metals concentrate in the fly ash, the flyash needs to be conditioned with water as it is discharged from the silo into trucks for delivery to the landfill. Within 24 h the conditioned flyash hardens into a concrete-like material which, although not high-strength, does not absorb water. Testing of the

conditioned flyash using state-mandated procedures indicates that the heavy metals are permanently bound and will not leach.

The SEMASS facility is provided with state-of-the-art acid-gas scrubbers [24]. The first two boiler units have electrostatic precipitators (ESPs) of a size greatly increased compared with previous units. The third boiler was installed using fabric filter emission controls. This facility has emissions well below US EPA standards, and even below standards of the European Union (see Table 3).

Palm Beach Florida Resource Recovery Facility

Babcock and Wilcox designed the boilers used for dedicated combustion of RDF produced in Biddeford, Maine, and Oronton, Maine, in 1987 and 1988, and later in Palm Beach, Florida, in 1989. The Palm Beach facility is another example of large-scale successful operation of an RDF combustion facility.

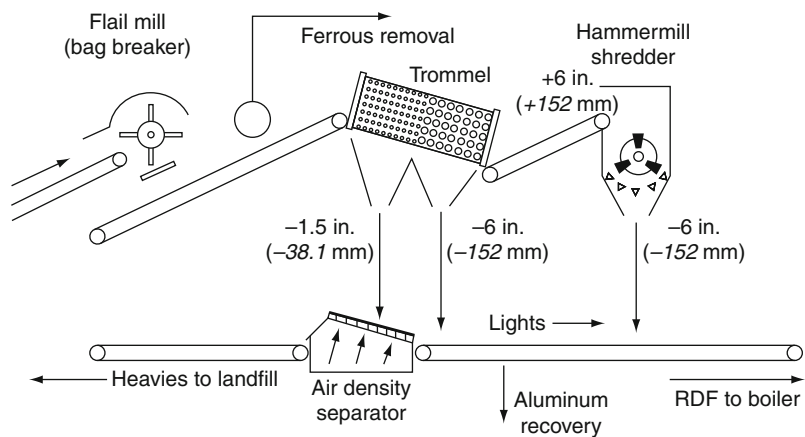
A model of the complete facility in Palm Beach is shown in Fig. 8. Municipal solid waste trucks enter the tipping floor and dump their loads on the floor. Front-loaders move the waste either directly to a metal pan feed conveyor, or to storage on the floor for later feeding. The loader operators inspect the waste as it is received, diverting over-size objects to a separate area, for recycling or disposal to landfill. The loaders feed the waste materials continuously, one bucket-load at a time, and deposit it onto feed conveyors. The front-loader operators in effect feed the conveyor as empty conveyor space appears, so that the feed is essentially steady. Shredders operate most efficiently and effectively at a full loading of the drive motor. Typically shredders are provided with 1,000 hp motors. The operators in the control room control the speed of these conveyors, so that the rate of feed is adjusted to the capacity of the shredders. Storage capacity is provided on the tipping floor as needed to handle the waste received at peak rates during the day while it is removed for processing at a steady rate.

The Palm Beach Facility has three shredders serving two basic processing lines. The feed conveyors are located in open space, rising to the shredder feed opening, and taking material out of the bottom of the shredder, again rising to feed the trommels, and again rising to feed the trommel screens. The RDF is then



Waste-to-Energy using Refuse-Derived Fuel. Figure 8

Palm Beach resource recovery facility [13]



Waste-to-Energy using Refuse-Derived Fuel. Figure 9

Processing with flail mill before screening and shredding, Palm Beach [13]

carried by conveyors directly to the boilers or bypassed to a storage floor for future feeding to the boilers. Storage capacity is needed so that the boilers can operate 24 h/day, weekdays and weekends, and also to allow for shutdown of boilers, preferably one at a time.

RDF is produced using two-stage shredding and trommel screens, as shown in Fig. 9.

A flail mill is used to break the bags of refuse, avoiding excessive grinding of glass. A magnetic separator placed above a conveyor removes ferrous material,

especially cans, which are baled for shipment to metals recovery plants. The coarsely shredded waste is then screened by a trommel in two stages; the minus-1-1/2 fraction which contains broken glass, dirt, and miscellaneous fines, drops to an air density separator which blows light materials such as paper off the surface while the heavies pass through the screen and are removed as heavies to the landfill. Minus-6-in. holes pass aluminum and remaining steel cans, which are baled for shipment to recovery plants. The plus-6-in. material is then fed to a hammer mill shredder, which grinds the paper, cardboard, plastics and wood. The last operation is the disk screen that passes minus-6 in. material joining the screened lights to become the “prepared fuel” or RDF product that is conveyed to the storage room for feeding to the boilers.

Three shredding lines serve the plant, two having a capacity of 100 t/h. The raw MSW enters the flail mill, and passes under a belt magnet. The collected ferrous metal is then shredded, passing a second magnet that returns the entrained matter to the conveyor, while the ferrous material is air scrubbed to produce 5 t/h of clean ferrous material.

The two-stage trommel drops the minus 1.5-in. material to an air density separator to produce 10 t/h of residue that goes to landfill. In this screening operation about 1.3 t/h of moisture is lost to the atmosphere. The RDF from the air density separator, air classifier, and disk screen is collected to produce an output of 83 t/h that is fed either directly to the boilers or stored on the tipping room floor. Typically, 0.7 t/h of aluminum is recovered. A shear shredder is shown, reducing tires to a size that can be burned in the boilers along with the RDF. The trommel removes the plus-2 in. material that is recycled to the shear shredder (Fig. 10).

The RDF is stored in a separate building, as shown in Fig. 11, from which it is loaded by front-loaders to the conveyors feeding the boilers. Shown are the RDF feeders that deliver to each of the boilers at speeds controlled by the operators in the control room. Figure 11 shows the boiler, the dry scrubbers and fabric filters followed by the induced-draft fans which discharge the combustion gases to the stacks.

The RDF feed system in front of the boiler is shown in Fig. 11. The feed conveyor delivers the RDF to the

feed bin. A ram feeder pushes the feed onto the feed conveyor, which in turn carries the feed up and over to the drop chute after which it feeds the air-swept spout which injects it into the boiler (Fig. 12).

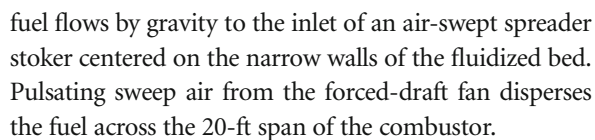
The air-swept spout is shown in Fig. 13. The RDF is directed over the traveling grate stoker upon which some of it is burned directly, and above which much of it is burned in suspension.

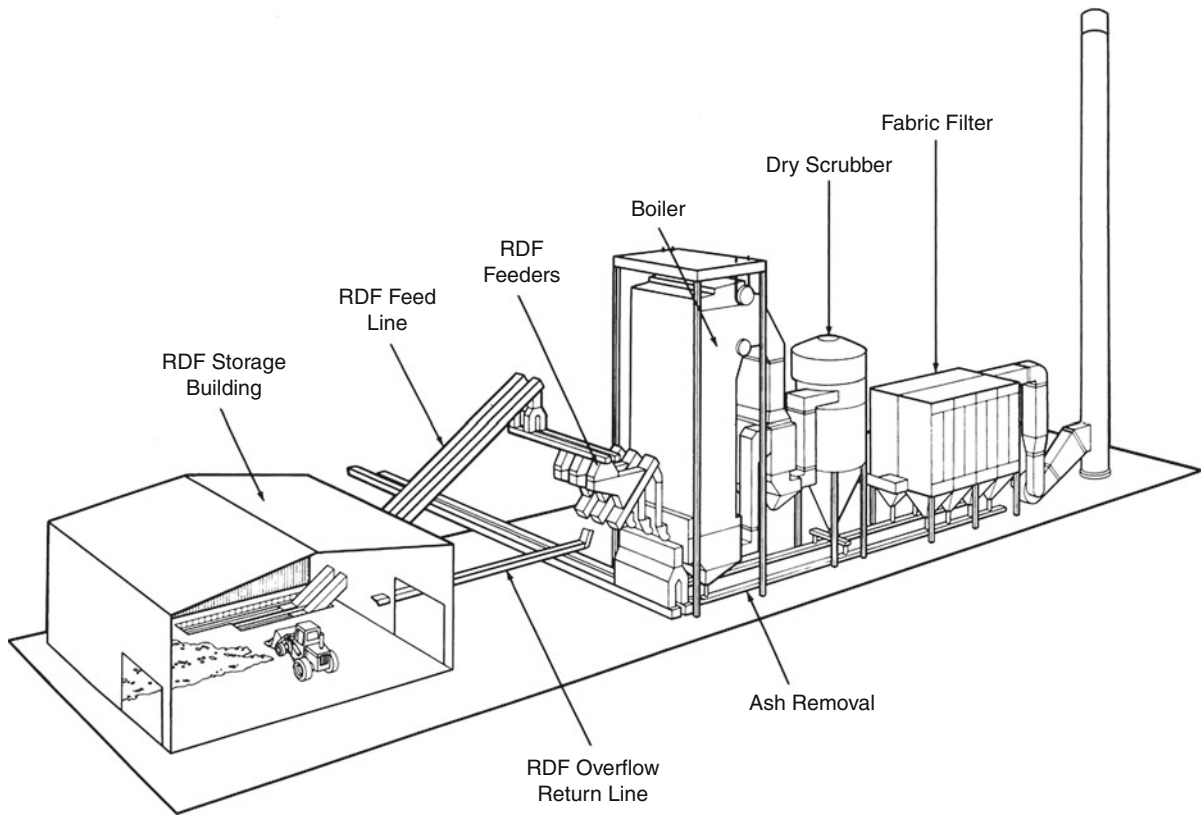
The ash residues are carried back to the ash hopper. Under-grate air supplies a portion of the combustion air. Secondary air (not shown) is supplied above the stoker.

Figure 13 shows the design of the Babcock and Wilcox combustion chamber above the firing grate. Overfire air is supplied at high pressure and passes through nozzles that direct the air so that turbulence is created in the combustion space, achieving effective mixing of fuel and air. The furnace walls are composed of alloy-faced or composite tubing that resists the intense heat and radiation and delivers the heat to the boiling water, to be released as steam. Figure 14 shows tubing which is used by Babcock and Wilcox, wherein the carbon steel core tubes are coated with stainless alloy in their manufacturing process. Some boiler manufacturers use solid alloy tubes, or apply the heat-resistant alloy coating after manufacture of the panels of tubes, and replace or restore the coatings periodically during annual shutdowns.

Ravenna, Italy Fluidized Bed System

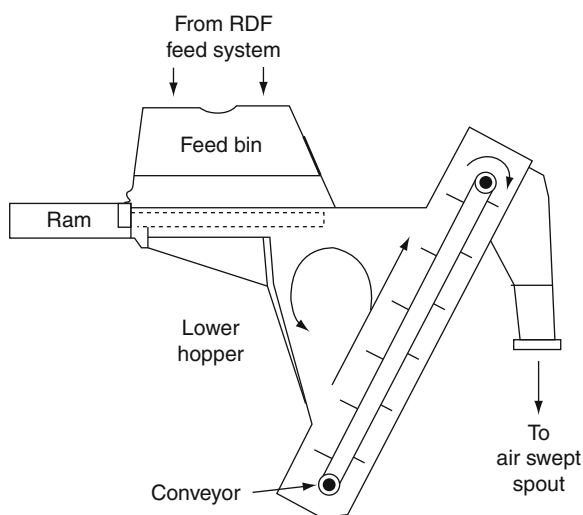
The City of Ravenna, Italy, a seaport on the Adriatic, in 1998 commissioned a fluidized bed boiler/waste-to-energy system to combust approximately 50,000 t/year of RDF and generate electric power. The chosen site for a plant to convert this waste to energy is close to a landfill, whereas the RDF production plant is 26 km away. It was decided to pelletize the RDF to facilitate transportation by truck. In order to employ a pelletizer, it is necessary to process the MSW by shredding and screening the product. The processing of waste includes material sizing by shredding to size, followed by magnetic separation. The final waste stream is compressed through roller mill/punched grate to create a relatively “densified” fuel stream. The fluidized bed boiler system was designed to deliver about 70,000 lb/h of steam to generate approximately 7 MW of electricity [13].





Waste-to-Energy using Refuse-Derived Fuel. Figure 11

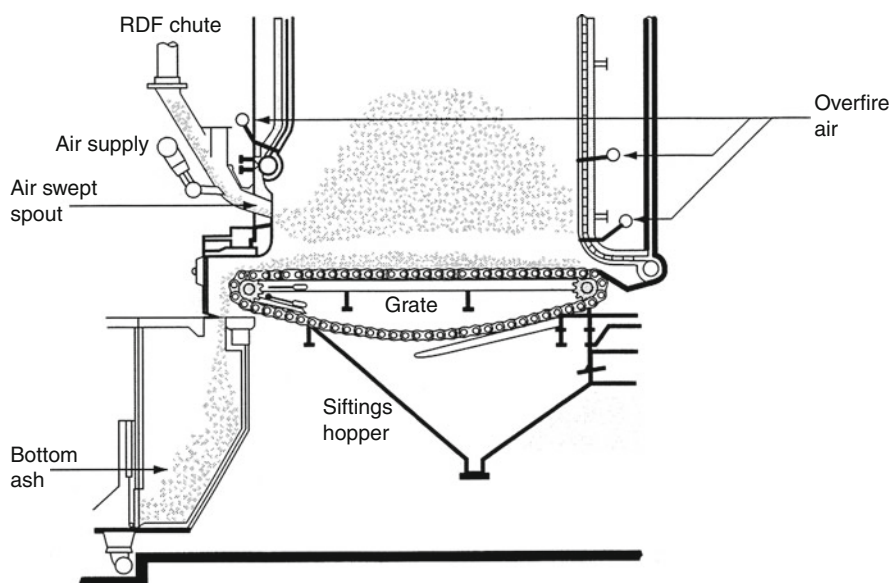
Perspective drawing of complete RDF system with dry scrubber and fabric filter



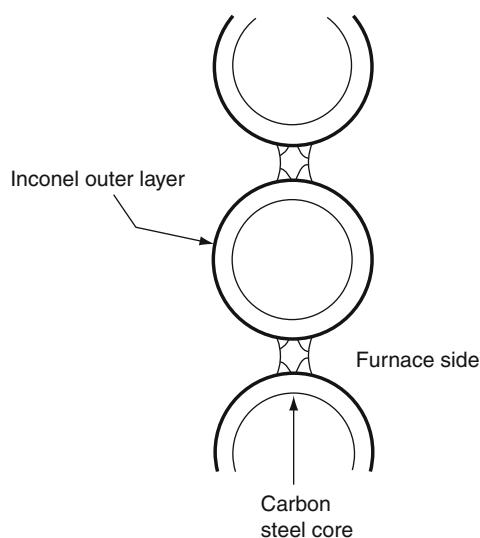
Waste-to-Energy using Refuse-Derived Fuel. Figure 12
RDF feeding system [13]

The fluid bed combustor is comprised of an internally insulated and refractor lined chamber having internal dimensions of approximately $10' \times 20'$ and a height of approximately 42 ft. The bottom of the combustor is equipped with a pair of draw-down cones that comprise the patented and proprietary bed cleaning system that is the trademark of EA Technology. Fluidizing air from the FD fan is distributed throughout the combustor by a series of manifolds and nozzles that distribute the air uniformly. In addition to the underfire air, a pair of overfire air manifolds supplies and distributes the balance of the combustion air into the lower region of the vapor space. The 52 ft elevation above the fluidizing nozzle provide 4 s of residence time within the chamber with approximately 2.5 s of that after the second level of overfire air injections.

The furnace is designed to operate at about 50% excess air, sufficient to control combustor temperature



Waste-to-Energy using Refuse-Derived Fuel. Figure 13
Elevation of lower section of boiler [13]



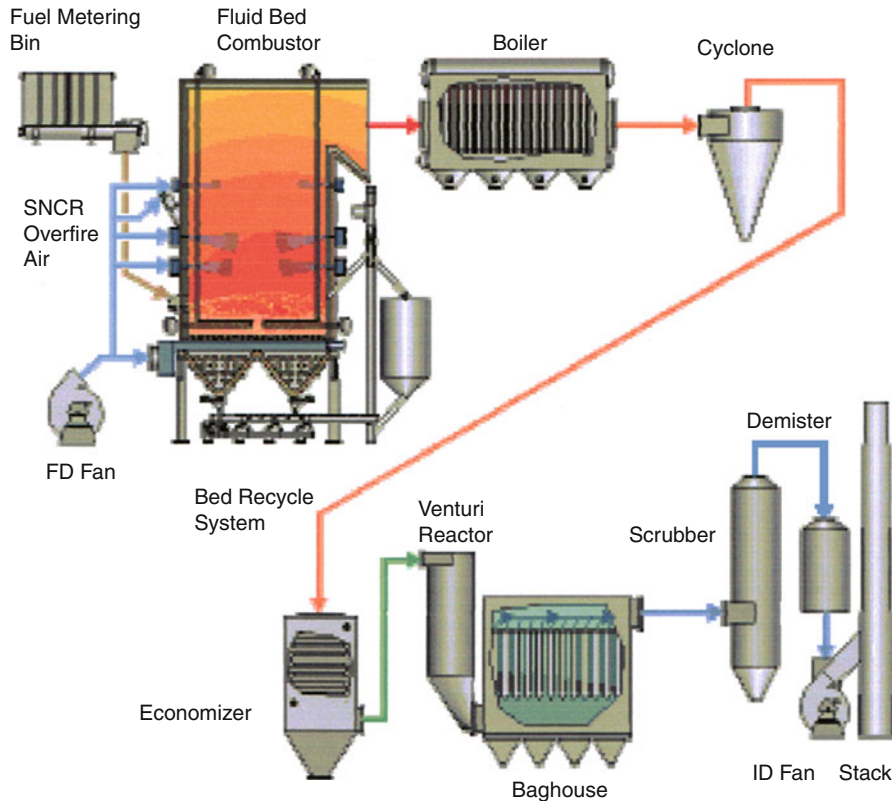
Waste-to-Energy using Refuse-Derived Fuel. Figure 14
Cross section of Babcock and Wilcox boiler tubes in waterwalls [13]

suppression: heat removal tubes provide the additional cooling effect for the gases to exit the furnace at about 1,700°F before entering the convective screen section where they are reduced to nearly 1,300°F before entering

the superheater. Further steam production occurs in the final evaporator section designed to reduce the flue gas temperature out the boiler to 650°F (342°C).

Exiting the boiler the flue gases are cleaned of some particulate by a set of multicyclones. After the economizer the gases are reduced to about 325°F (162°C) while heating feedwater to around 390°F (199°C). The boiler is designed to generate 75,000 lb/h of steam at 595 psia and 716°F (41 bars, 380°C).

Emission Control Downstream of the boiler, particulate levels are controlled by a venturi reactor and bag house. The venturi reactor is a dry scrubber that incorporates a solid recirculation system, which optimizes the reactant equalization ratio. Hydrated lime and activated carbon are added as needed to reduce the acid gases and minimize dioxin/furan emissions. The flue gases at 300°F then enter a wet scrubber where they are quenched to 130°F. A solution of NaOH is sprayed into the flue gas prior to a series of demisters for final removal of any remaining acid gases and other soluble components of the gases. Prior to exiting the system through the induced-draft fan, the exhaust gases are reheated to 239°F (115°C). This comprehensive



Waste-to-Energy using Refuse-Derived Fuel. Figure 15

Process flow schematic of Ravenna fluidized bed system firing refuse-derived fuel [15]

emission control system reduces emission levels well below the already stringent requirements of the Italian government [15].

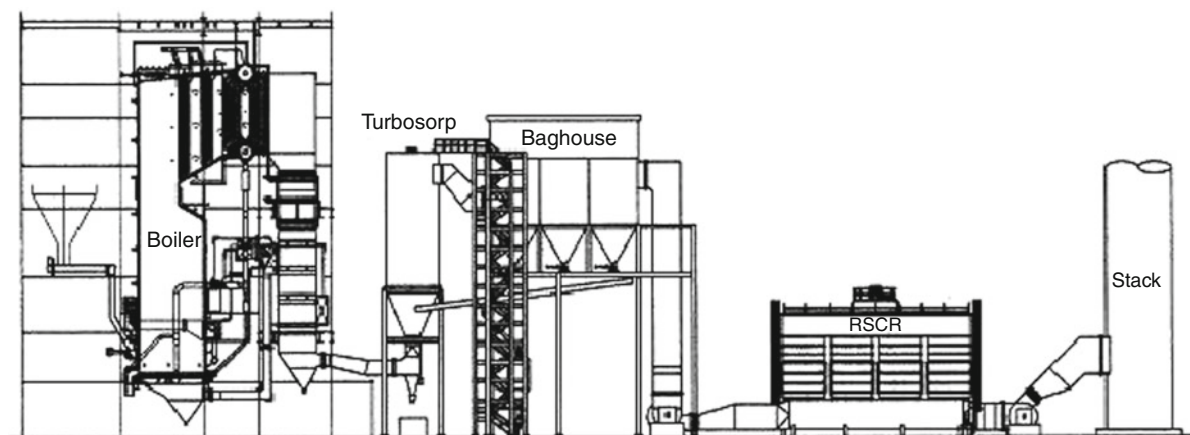
Alliance, Ohio RDF Facility

A “greenfield” RDF facility in Alliance, Ohio, will process 2,400 t/day (TPD) of MSW and Construction and Demolition Debris (C&D). Two equipment trains to thermally process the material will consist of Riley Power’s Advanced Stoker boiler, Turbosorb dry scrubber and the Regenerative Selective Catalytic Reduction (RSCR) nitrogen oxides (NO_x) control system. The boiler/emission control system is shown in Fig. 16 [16].

The Riley Advanced Stoker boiler has unique design features to ensure high efficiency, corrosion resistance, and fuel flexibility. The use of the Turbosorb will result in lower emissions of lead and other volatile heavy metals and mercury than a typical spray

dryer/baghouse (Spray Dryer Absorber) system can achieve. Acid gas removal is also superior to an SDA system while utilizing less lime reagent and power. A Riley Selective Catalyst Remover (RSCR) follows the Turbosorb and baghouse as a “low-dust” Selective Catalytic Remover, having an auxiliary energy consumption about 85% lower than a typical low-dust, tail-end SCR. The RSCR reduces nitrogen oxides (NO_x) and carbon monoxide (CO) emissions to lower values compared with previous facilities producing energy from waste. This versatile system is capable of controlling emissions from combustion of commercial and demolition wastes (C&D) as well as refuse-derived fuel (RDF) from municipal waste.

The RDF and C&D conveyed to two Riley boilers and associated steam turbines will produce a total of 66 MW gross, which results in approximately 58 MW net electric power. At maximum continuous rating (MCR), 83,333 lb/h of RDF and C&D will be sent to each boiler.



Waste-to-Energy using Refuse-Derived Fuel. Figure 16
Alliance Ohio boiler and AQC equipment [16]

The design higher heating value of the fuels is 4,500 BTU/lb for the RDF and 5,681 BTU/lb for the C&D, combined at 50%.

Figure 17 shows the Riley Advanced Stoker Boiler with a modular boiler bank, which replaces the conventional boiler bank. This design allows the evaporators to be shop assembled, resulting in fewer field execution hours.

A Turbosorp Dry Circulating Fluidized-Bed Scrubber (Fig. 18) removes acid gases from the boiler flue gas using lime injection. It is followed by a fabric filter (baghouse), which controls particulate emissions. Activated carbon injection is used to remove heavy metals, including mercury, as well as dioxins/furans. The flue gases enter a central pipe, rise to the top and then fall down through the annular outer pipe. This arrangement is very compact compared with the spray-dry scrubbers commonly in use.

The benefits of the Turbosorb technology over the spray dryer absorber (SDA) used heretofore include:

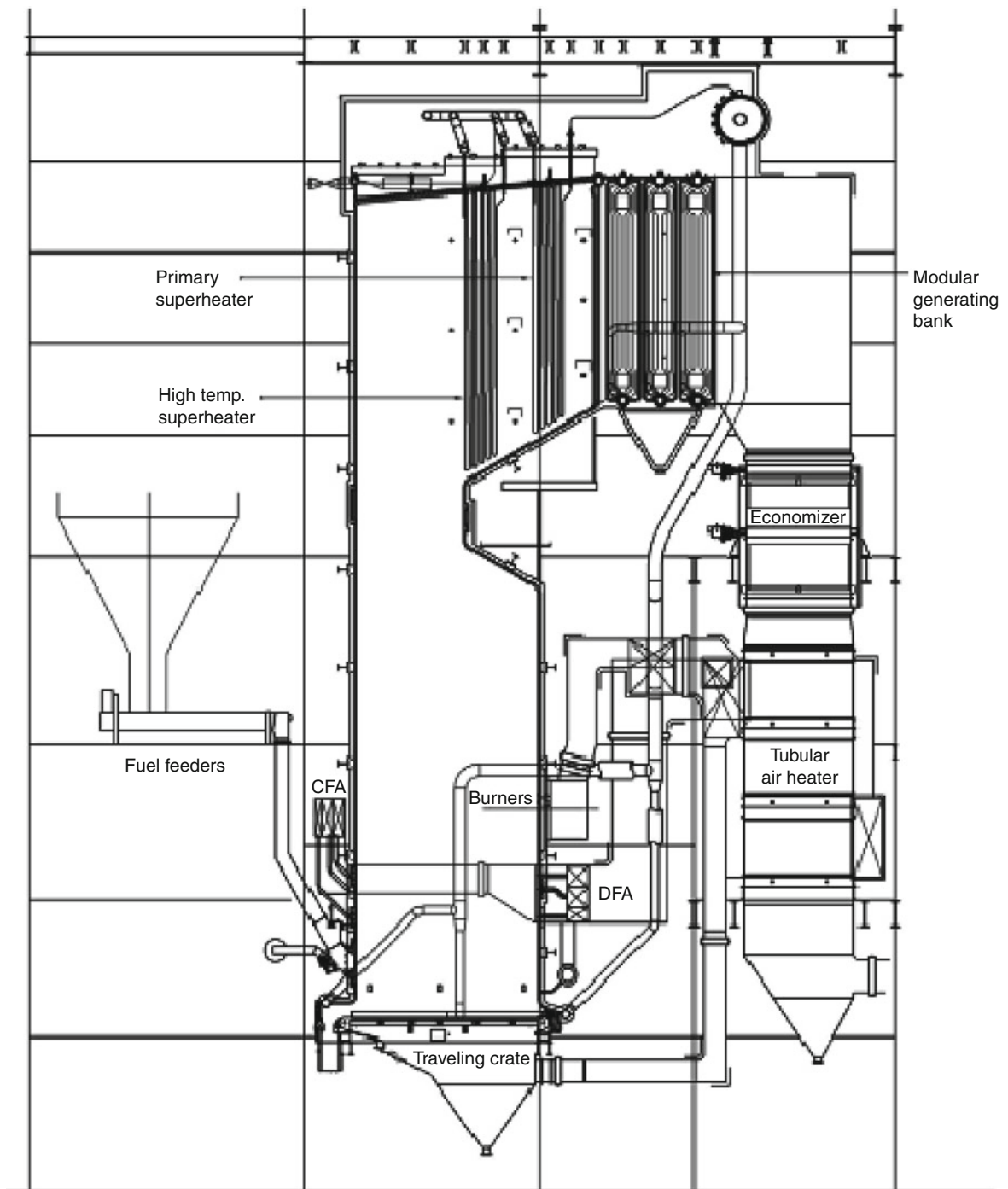
- Lower lime stoichiometry for the same removal efficiency and approach temperature.
- Higher removal efficiencies for acid gases, i.e., 95% or greater plus removal of sulfur trioxide (SO_3) sulfuric acid (H_2SO_4), hydrochloric acid (HCl) and hydrofluoric acid (HF), mercury, particulates, and dioxin/furans.
- Reduced power consumption since slurry atomization is not used.

- Simpler system with less maintenance since it uses dry lime, without the need for slurry handling.

The Turbosorp dry scrubbing technology employs high levels of solid circulation, with addition of finely atomized water, hydrated lime, and flue gas within the circulating bed reactor. The lime and finely atomized water are injected independently into the Turboreactor to lower flue gas temperature and enhance absorption capacity. The fluidized bed material is comprised of solids, including calcium hydroxide, recirculated flyash from the combustion process, and solid reaction products for the fabric filter.

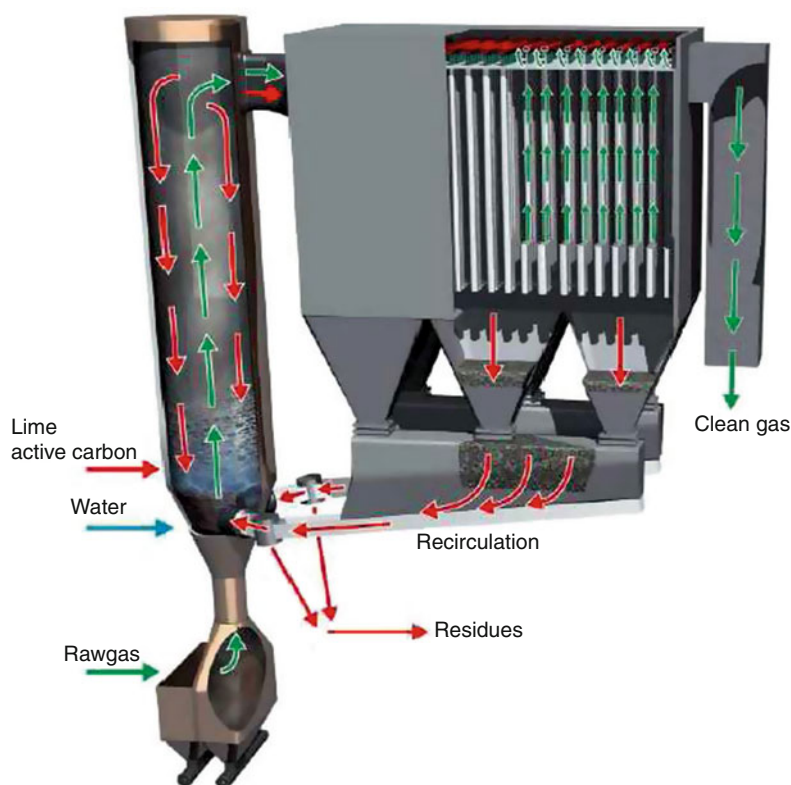
Turbulent flow within the reactor permits intimate contact between the gas and the reagents promoting the chemical reaction between the gas and the active portions of the reagent. A portion of the solids is drawn out of the system and recirculated to maintain mass equilibrium. Provision to inject activated carbon into the gas stream will be made, but may not be required to achieve the emissions limits.

The Regenerative Selective Catalytic Reduction (RSCR) system shown in Fig. 19 is an advanced high-efficiency NO_x reduction system. Since the flue gas temperature is too low for the SCR catalyst to function efficiently, the RSCR temporarily elevates the flue gas temperature to the optimum level while using minimal auxiliary fuel, allowing NO_x reduction efficiencies greater than 70% to be achieved.



Waste-to-Energy using Refuse-Derived Fuel. Figure 17

Riley Power Inc. advanced stoker boiler [17]



Waste-to-Energy using Refuse-Derived Fuel. Figure 18
Turbosorp and fabric filter filtration gas scrubbing system [17]

The RSCR device is a multi-chamber vessel, each chamber containing catalyst and ceramic beds [17]. Each chamber is arranged for vertical flow of gas in both the upward and downward direction. Dampers in the gas ductwork control the direction of gas flow. The catalyst mounted directly above the ceramic beds promotes the reduction of NO_x to nitrogen and water, while the ceramic bed is used to recover heat from the processed flue gas. The active reagent used for this process is 19% aqueous ammonia.

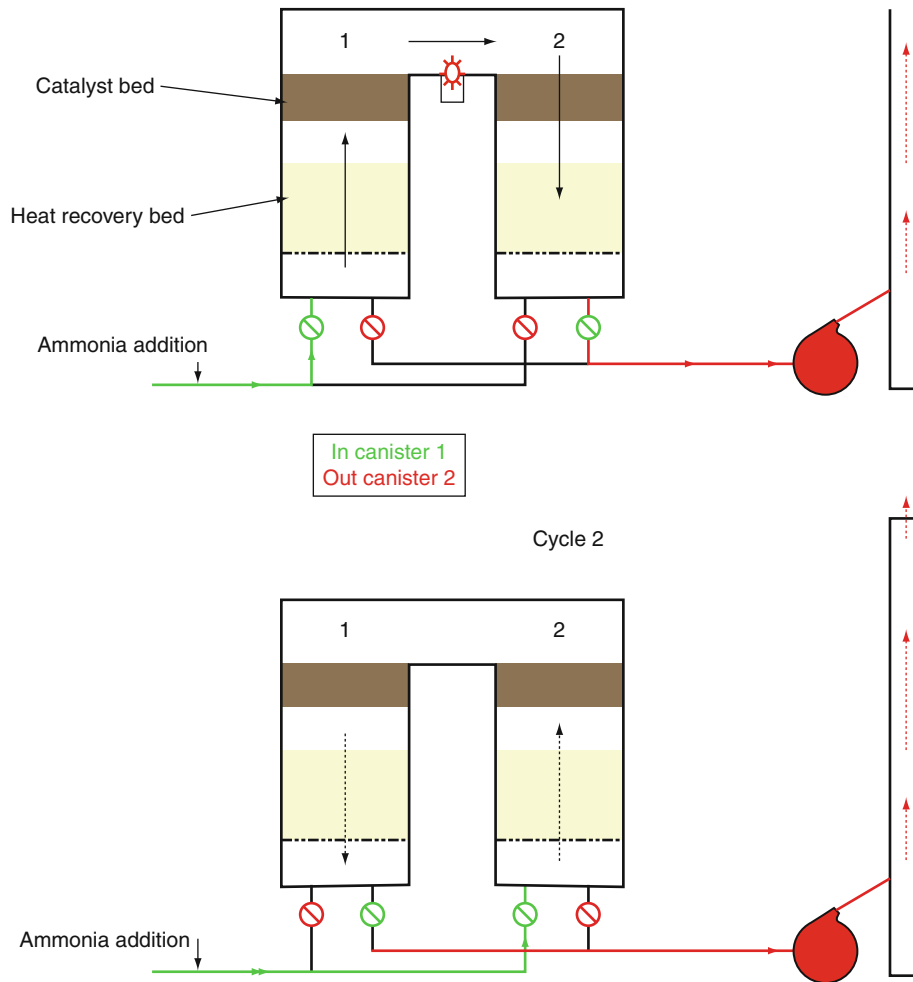
Flue gas enters an inlet header that connects to the four reaction canisters. At any time, one inlet duct and one outlet duct on each pair of canisters is open, and one of each is closed. Flue gas enters one canister and rises through a heat recovery medium, picking up heat to attain the proper temperature for NO_x removal (about 600°F). It then enters the catalyst, causing NO_x to be converted to nitrogen and water. After passing through the first catalyst, a small quantity of natural gas is combusted to raise the temperature of the flue gas by

about 10°F. The gas then turns and goes down through the second canister, first passing through the catalyst, causing additional NO_x to be removed, then through the heat recovery medium, where the heat in the flue gas is transferred back to the medium. The gas then exits through the outlet header to the stack.

About every 3 min, the inlet and outlet dampers switch positions, reversing the flow. The flue gas picks up heat from the now heated side and gives up the heat to the cooler side. This process cools the gas to approximately 280°F before exiting to the stack.

Fairfield Renewable Energy Power Plant

Energy Answers International is developing a renewable energy power plant on an industrial-zoned site on Fairfield Peninsula in Baltimore, Maryland. The 140 MW combined heat and power plant is designed to provide wholesale energy to help meet regional demands, as well as reduced-price retail energy as an attraction for energy



Waste-to-Energy using Refuse-Derived Fuel. Figure 19
Regenerative selective catalytic reduction (RSCR) nitrogen oxides (NO_x) system [18]

intensive industries to co-locate at an Eco-Industrial Park to be developed on the 90-acre “brownfield” site. The waste-derived fuel will be processed at various other sites and transported to the plant.

Processing Equipment

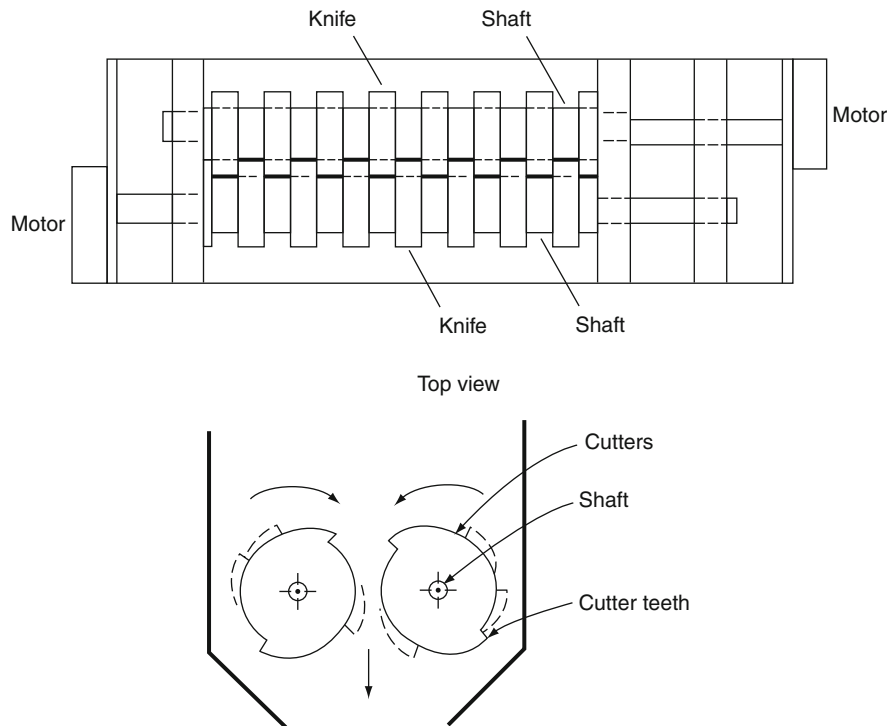
Typical equipment, common to RDF production, is described below.

Shredders for Size Reduction

Figures 20–22 show the typical shredders employed in RDF processing: shear shredders, flail mills, and hammer mills [1].

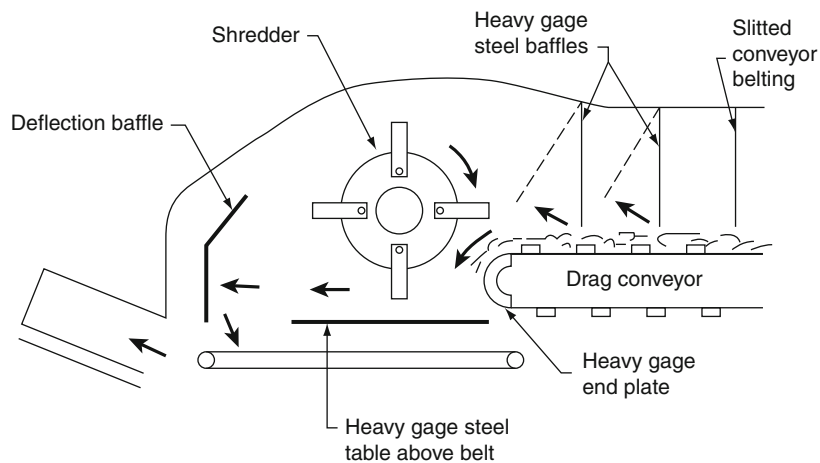
The flail mill has simple swinging hammers that reduce the size of bulky boxes, cardboard, and other materials, breaking bottles into large shards as the material is smashed onto the heavy steel table.

The shear shredder consists of counter-rotating wheels having knives or cutters that shear the materials as they are caught in the gap between the wheels. Hydraulic drives are used, so that when jams occur, the rotor can be stopped and started again to complete the grinding process. These machines are especially suited to processing tires, but also to screening RDF. The spacing of the wheels and the gap between the drums determines the size of the product.



Waste-to-Energy using Refuse-Derived Fuel. Figure 20

Shear shredder

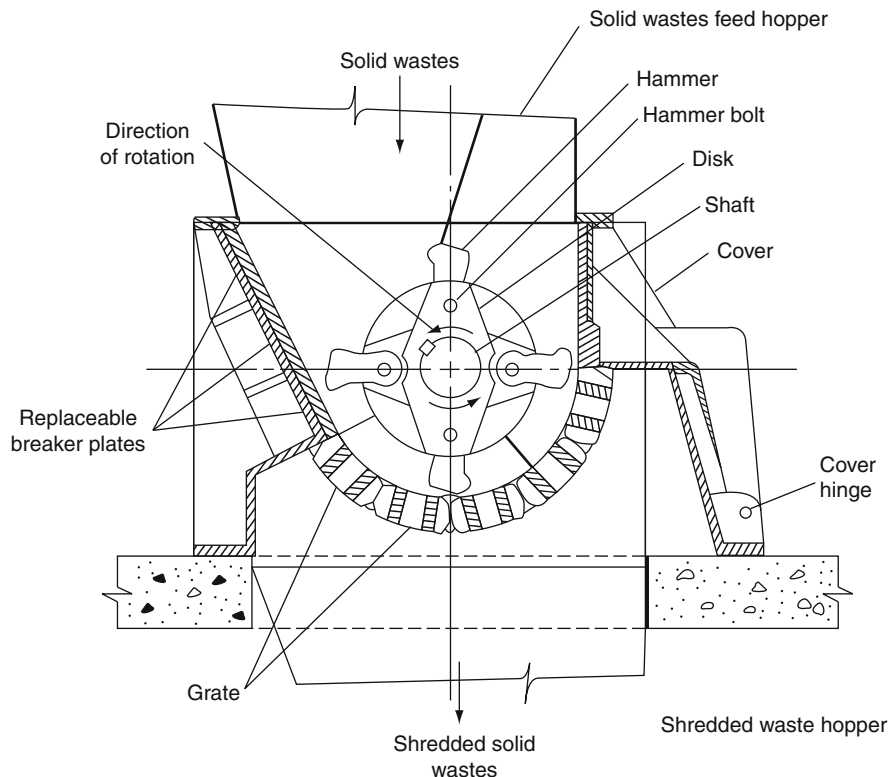


Waste-to-Energy using Refuse-Derived Fuel. Figure 21

Flail mill

A typical hammer mill has hammers mounted on the shaft, free to swing as they pass over grates on a quadrant of the circular shell. The hammers break the material by impact, and then force it through the grates as it is reduced in size. Material that is not

shredded on the first pass circulates around until it can pass the grates. The spacing of the grates determines the size of the product. The rate of wear on the hammers is such that they have to be reversed or replaced periodically, typically on weekend downtime.



Waste-to-Energy using Refuse-Derived Fuel. Figure 22
Hammer mill

Magnets for Ferrous Recovery

Magnets are universally employed to separate ferrous metals from a stream of shredded refuse.

Figure 23 shows a typical Eriez magnet suspended above a belt conveyor. The magnet has a transverse belt that carries the collected metal away from the conveyor to a collection bin. An air density separator is employed to clean off entrained light material and return it to the RDF conveyor.

Screens Used in Resource Recovery

A variety of screens are used in RDF processing as shown in Fig. 24 – (a) vibrating flat, (b) vibrating double flat, (c) inclined flat, (d) disk, (e and f) rotary or trommel. Screens typically separate materials of smaller size from the mass passing over the screen, the size determined by the diameter of the hole, or the area of the passage. Rotary screens, called trommels, are especially suited to the operation, since the material is

tumbled over the screen, exposing new material on every revolution.

Flat screens generally provide a vibratory or shaking motion, to assist the material in falling through the apertures in the screen. A flat screen, which incorporates an upward air flow, is shown in Fig. 25. This provides effective separation of light material from the heavy residue.

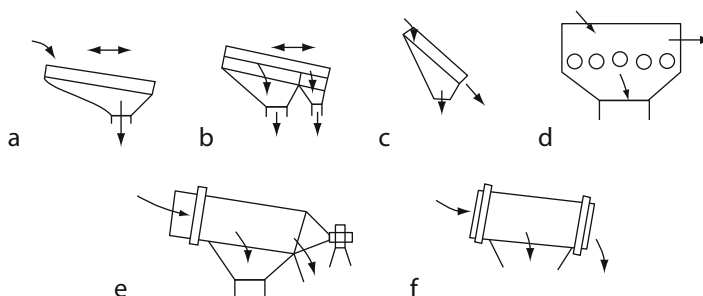
The Jeffrey Rader screen shown in Fig. 26 has rotating wheels provided with cutters: the spacing of the cutter drums and the gap between the opposing drums determines the size of the material that can pass through the screen.

Conveyors for Movement of Materials

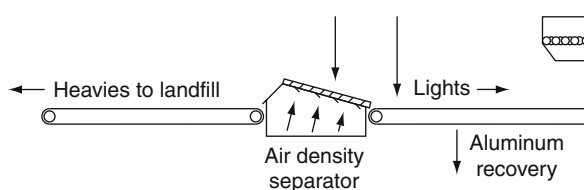
Belt conveyors are the most common devices used for carrying RDF from one operation to another. The belts are carried on idler rolls, and are driven by a head pulley. Drop-off of material from one conveyor to



Waste-to-Energy using Refuse-Derived Fuel. Figure 23
Horizontal magnet used for ferrous recovery (Eriez)



Waste-to-Energy using Refuse-Derived Fuel. Figure 24
Screens



Waste-to-Energy using Refuse-Derived Fuel. Figure 25
Flat screen with upward air flow

another is a source of dusting and linting within the processing areas. Suitable covers must be used, and often a dust-collecting system is provided, which returns the dust to the processing line.

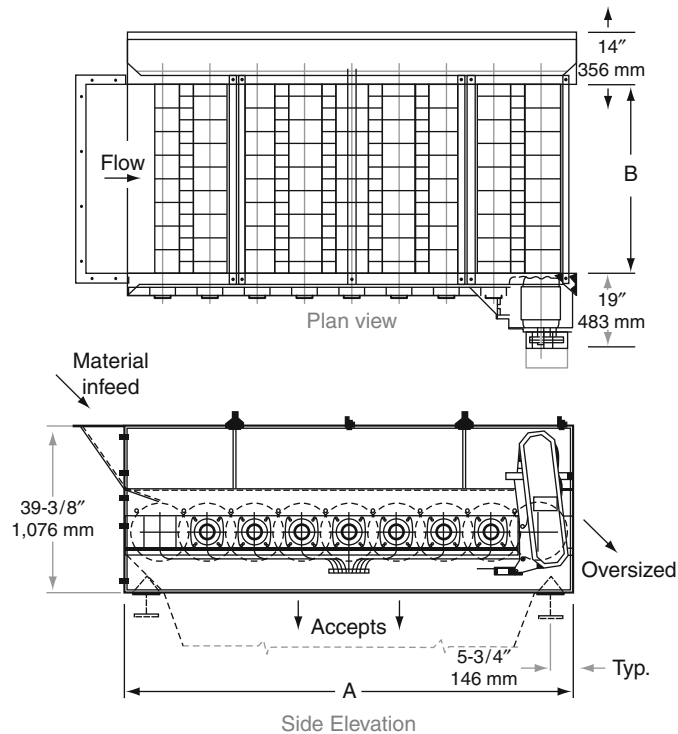
Belt conveyors are generally troughed to minimize fall-off of material. The slope of belt conveyors is

limited to about 30° , so that material will not fall back (Fig. 27).

Vibrating-pan conveyors are employed with heavy or dense residual materials, including broken glass and combustion ash residues. Figure 28 shows the principles of operation. The movement "B" causes a rise and fall of the pan. The particles are suspended and fall back in a different place on the pan, progressing the material along. The oscillating drive forces movement over isolation springs, which are tuned to the natural frequency of the pans.

Steel Pan Conveyors

Pan conveyors as seen in Fig. 29 are used to handle raw municipal waste and other solid materials. The pans are hinged, and connected to steel rollers carried on tracks.



Waste-to-Energy using Refuse-Derived Fuel. Figure 26

Rotating disk screen for separating materials by size (Jeffrey Rader)



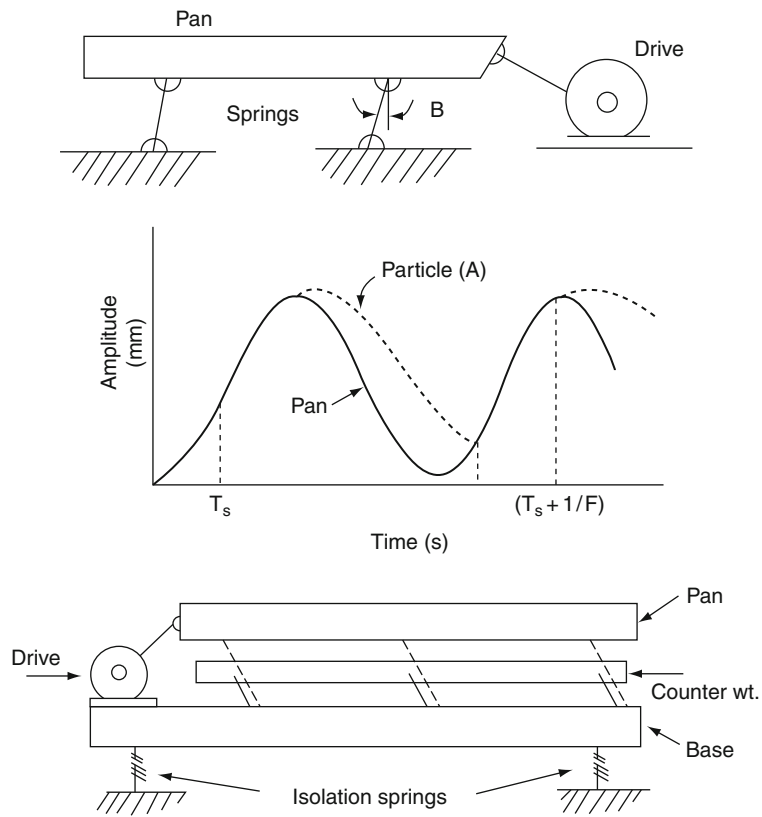
Waste-to-Energy using Refuse-Derived Fuel. Figure 27

Belt conveyor for moving light materials (typical)

Air Quality Control Equipment

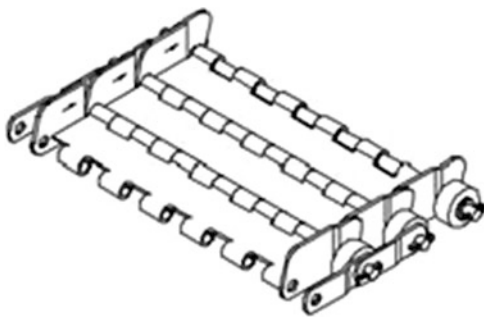
Early Systems with Electrostatic Precipitators

The complete boiler and emission control system typical of early installations is shown in [Fig. 30](#). A feed cyclone is shown, which drops the RDF into the air-swept spout onto the stoker. The boiler furnace has the rectangular shape commonly used in coal and wood stokers at the time. A “nose” deflects the gases into the top of the chamber where they enter the economizer. Next the gases pass through the convector section of the boiler where most of the steam is generated before passing through the superheater. An economizer is shown, cooling the gases before entering the electrostatic precipitator (ESP). The economizer heats the boiler feedwater before it enters the convection section



Waste-to-Energy using Refuse-Derived Fuel. Figure 28

Vibrating conveyors used to move heavy materials such as ash residues [1]



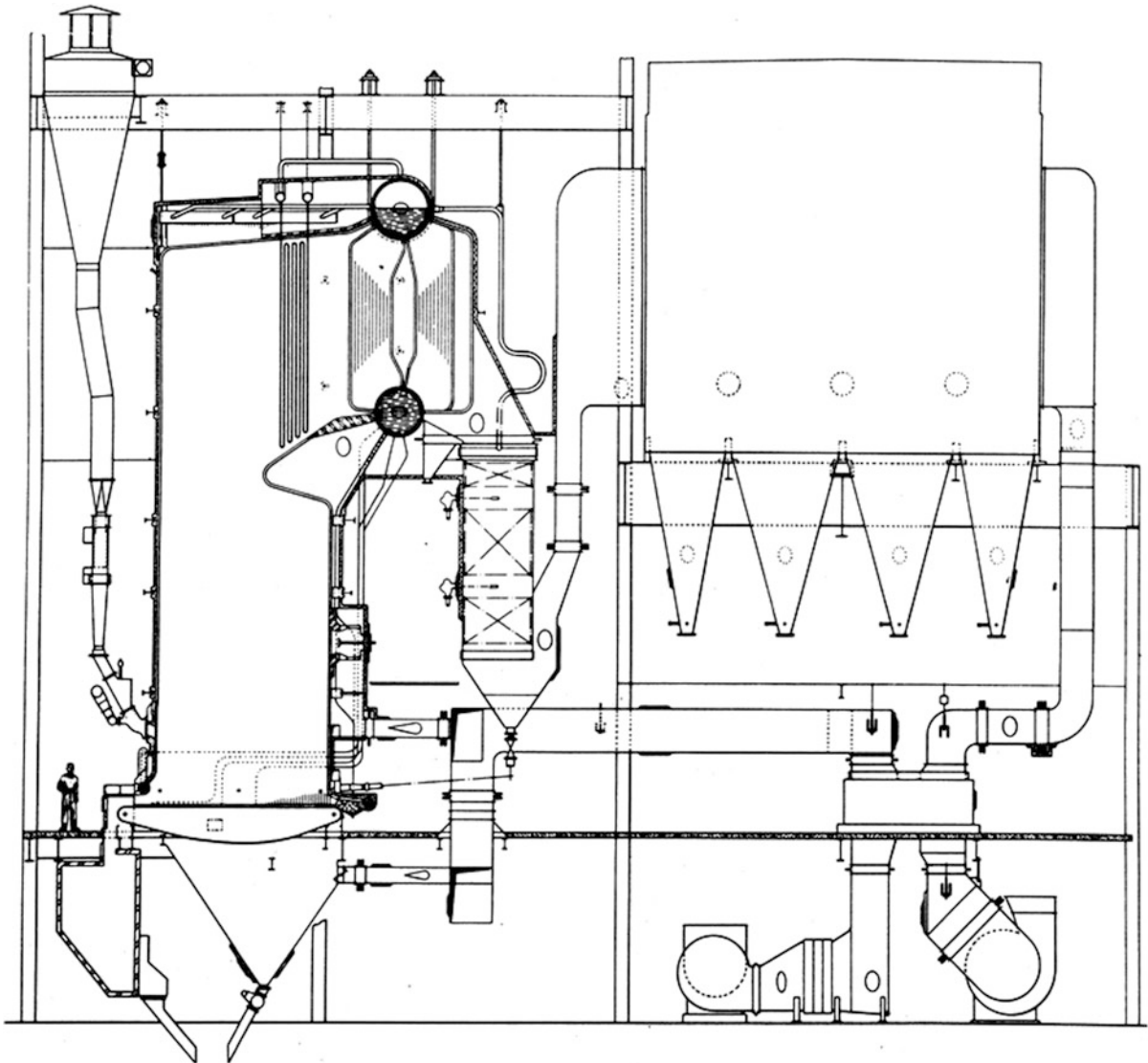
Waste-to-Energy using Refuse-Derived Fuel. Figure 29

Steel pan conveyor used for handling raw MSW

of the boiler. An air heater is shown, below the ESP. This preheats the combustion air, as needed to assist in driving off moisture in the RDF as it is fed to the stoker.

Contemporary Systems with Scrubber and Baghouse

Scrubbers and fabric filters, “baghouses,” have replaced electrostatic precipitators in contemporary waste-to-energy plants. Dry scrubbers are commonly used to capture sulfuric acid and hydrochloric acid from emissions from waste-burning systems. Figure 31 shows flue



Waste-to-Energy using Refuse-Derived Fuel. Figure 30
Early Babcock and Wilcox RDF-fired stoker boiler with electrostatic precipitator

gases entering a spray-dryer tower within which a solution of caustic solution of lime, prepared by dissolving calcium oxide (CaO) in water, is atomized and sprayed into the gas stream passing through the tower. The scrubber functions by using the temperature potential of the entering gases (typically above 400°F) to evaporate the water in the droplets of lime, resulting in drying of the droplets while absorbing the acids, leaving the spray dryer at a temperature of about 350°F . The lime has reacted with the acids, producing

calcium sulfate and calcium chloride as dry products that are collected by the fabric filter. Particulate matter (flyash or dust) entering with the combustion gases is also collected. Compressed air is used to loosen the dust from the surface of the fabric so the dry particulate and lime reactants will drop to the bottom of the filter housing, where they are released through a rotary sealing valve. This product must be “conditioned” by adding water in order to make “dust free” handling possible. The presence of free lime in the collected dust

helps to reduce or eliminate the toxicity of this product by forming compounds insoluble in water, reducing their environmental hazard.

Figure 32 shows the present technology of Babcock and Wilcox – a RDF-fired stoker boiler with dry

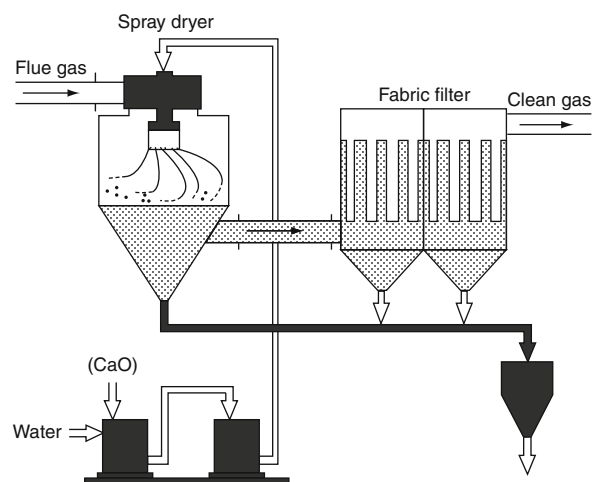
scrubber and fabric filter. The economizer and air heater are in a downward pass, and the gases enter the acid gas scrubber tangentially, leaving at the bottom as they enter the baghouse before the induced draft (I.D.) fan discharging to the stack.

Comparing Emissions of RDF Systems with those of Mass-Burn Facilities

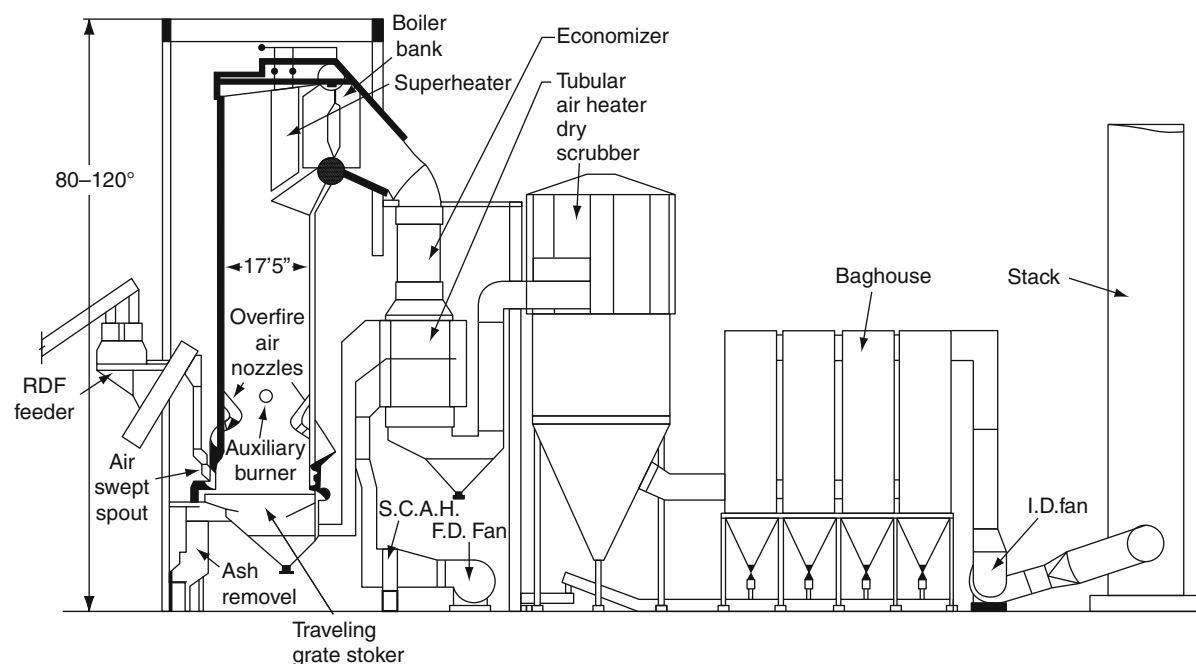
The final emissions that exit the stack are a direct result of the operational efficiencies of the project and the air quality control equipment in place. All waste combustion facilities can operate in compliance with permit limits by employing advanced control technology and operating efficiently.

A major difference related to air quality control equipment in a processed-fuel facility and a mass-burn incineration system is the amount of air and gases that are generated and therefore need to be cleaned.

A Mass-Burn incinerator requires significantly more air to combust the waste in the boiler. This air then must pass through the air quality control equipment. It has a larger air volume hence requires larger



Waste-to-Energy using Refuse-Derived Fuel. Figure 31
Spray-dry acid gas absorber tower



Waste-to-Energy using Refuse-Derived Fuel. Figure 32
Babcock and Wilcox RDF-fired stoker boiler with dry scrubber and fabric filter [13]

equipment with greater cost. Considering that the air quality control equipment is one of the largest expenses related to either technology, this can be a significant consideration in the technology choice.

Table 2 compares air emissions data from RDF and Mass-Burn combustion facilities. The RDF facilities show generally lower emissions of nitrogen oxides (NO_x) than do the typical mass-burn facilities, probably the result of more effective control of secondary air resulting in lower peak combustion temperatures. Carbon monoxide readings of the RDF facilities are higher, but this does not mean that trace organic emissions are high, as indicated by the extremely low PCDD/PCDF (dioxin) emissions measured. These were well below the European Union standard for dioxins. The mass-burn facilities as well as the RDF facilities today routinely employ activated carbon injection before the fabric filter to remove mercury, resulting in removing the dioxins before the gases leave the stack.

Table 3 shows the air quality test results for Boiler #3 at the SEMASS Resource Recovery Facility which has a scrubber and baghouse. The SEMASS current permit limits and MACT standards are provided for comparison.

In all tests, actual emissions were far below SEMASS permit limits, and the current MACT standards. . Note that these are the standards set forth in 40 CFR Part 60, Subpart Eb for new RDF-fired Municipal Waste Combustors. All emission standards and test results are corrected to 7% O_2 , dry basis, and standard conditions (68°F and 1 atm).

Properties of Municipal Waste and Processed Refuse Fuel

The organic content of these fuels is “renewable” in the sense that most of it is produced from organic sources, trees, food, gas, and oil. Placing wastes in a landfill is not “renewable” but rather represents a waste of resources that could replace oil, coal, and gas in the production of energy for heating and for power production.

Energy from the sun is renewable, and sustainable. Likewise, organic material grown under the sun is renewable. Once this material is grown, it is wasteful not to use it, as distinguished from use of non-sustainable renewable fuels such as coal, oil, and gas recovered from the ground.

Waste-to-Energy using Refuse-Derived Fuel. Table 2 Comparison of air emissions from RDF and mass-burn facilities [13]

Pollutant	Facility Units @ 12% CO_2	RDF		Mass		
		MERC ^a	Mid-	Bridgeport ^c	Bristol ^d	Stanislaus ^d
		Biddeford ME	Conn ^b Hartford CT	Bridgeport CT	Bristol CT	Stanislaus Country CA
NO_x	ppmd _v	202	190	225	280	300 ^e
CO	ppmd _v	83	192	23	19	43
VOC	ppmd _v	<1	<1	<2	<2	4
SO_2	ppmd _v	5	4	9	13	3
HCL	ppmd _v	10	2	6	8	2
Particulate	gr/DSCF	0.008	0.006	0.001	0.005	0.004
PCDD/PCDF	ng/NM ³ (USEPA Toxic Eq.)	0.04	<0.03	0.015	0.09	0.06

^a[19]

^b[20]

^c[21]

^d[22]

^eStanislaus Country NO_x emission without thermal De NO_x system in operation

Waste-to-Energy using Refuse-Derived Fuel. Table 3 Air emission test results of SEMASS Boiler No. 3 (1994 through 2008) (©2010 Energy Answers International, Inc)

Regulated substance	SEMASS boiler no. 3 permit limits	Current MACT standards (a)	SEMASS boiler No. 3 test results				
			1994	1995	1996	1997	1998
Particulate (mg/Ncm)	21.0	15.3	3.50	1.40	0.350	0.875	3.50
Sulfur dioxide (ng/Ncm or % removal)	59 or 80%	61 or 80%	29.9	48.2	19.0	31.7	34.6
			87.4%	87.0%	Not available	89.4%	89%
Hydrogen chloride (mg/Ncm or % removal)	29 or 95%	29 or 95%	3.25	5.11	4.20	3.23	5.11
			96.2%	96.9%	99.5%	99.3	99%
Nitrogen oxides (mg/Ncm)	263	220	197	230	176	212	216
Carbon monoxide (mg/Ncm [24-h] or [97% of 10-min values])	134	134	55.5	5.0	56.6	46.1	41.9
Cadmium (µg/Ncm)	30.6	15.3	0.99	3.1	<0.084	0.29	0.326
Lead (µg/Ncm)	336	107	<18.6	<60.0	4.87	22.2	9.10
Mercury (µg/Ncm or % removal)	21.4	38.2	4.7	<6.7	2.81	2.64	2.62
	85%	85%	93.9%	Test not required	Test not required	Test not required	98.8%
Dioxins/furans (ng/Ncm [NATOTEQ] ng/Ncm [tetra-octa])	No standard	No standard	0.011	<0.015	0.00498	0.0122	0.0169
	22.9	9.9	0.586	<0.887	0.319	0.663	0.826
Regulated substance	SEMASS boiler no. 3 permit limits	Current MACT standards (a)	SEMASS boiler No. 3 test results				
			1999	2000	Jan 2001	Apr 2001	Apr 2004
Particulate (mg/Ncm)	21.0	15.3	3.15	9.27	1.15	1.19	5.9
Sulfur dioxide (ng/Ncm or % removal)	59 or 80%	61 or 80%	44.8	42.7	52.3	49.9	68
			80%	88%	78%	84.6%	85.6%
Hydrogen chloride (mg/Ncm or % removal)	29 or 95%	29 or 95%	6.96	5.22	11.5	9.85	5.69
			99%	99.3%	98.4%	98.7%	99%
Nitrogen oxides (mg/Ncm)	263	220	227	227	220	184	195

Waste-to-Energy using Refuse-Derived Fuel. Table 3 (Continued)

Regulated substance	SEMASS boiler no. 3 permit limits	Current MACT standards (a)	SEMASS boiler No. 3 test results				
			1999	2000	Jan 2001	Apr 2001	Apr 2004
Carbon monoxide (mg/Ncm [24-h] or [97% of 10-min values])	134	134	65.4	36.5	65.4	54.9	88
Cadmium (µg/Ncm)	30.6	15.3	0.634	0.520	0.34	0.28	0.72
Lead (µg/Ncm)	336	107	18.0	18.6	4.97	4.47	26.0
Mercury (µg/Ncm or% removal)	21.4	38.2	4.89	1.05	1.99	18.2	4.6
	85%	85%	94%	99%	99.0%	91.6%	96.8
Dioxins/furans (ng/Ncm [NATOTEQ] ng/Ncm [tetra-octa])	No standard	No standard	0.0864	0.013	0.0329	0.0138	0.0207
	22.9	9.9	5.27	0.910	2.53	1.00	1.29
Regulated substance	SEMASS boiler no. 3 permit limits	Current MACT standards (a)	SEMASS boiler No. 3 test results				
			Jan 2005	Nov 2005	July 2006	July 2007	Apr 2008
Particulate (mg/Ncm)	21.0	15.3	1.94	0.99	1.83	4.2	7.04
Sulfur dioxide (ng/Ncm or% removal)	59 or 80%	61 or 80%	57	52	77	16.0	29.1
			84.4%	81%	80.1%	80.1%	88.8%
Hydrogen chloride (mg/Ncm or% removal)	29 or 95%	29 or 95%	18	32	9	5.85	5.9
			97.7%	96.0%	99%	99.2%	99.1%
Nitrogen oxides (mg/Ncm)	263	220	188	235	219	135.8	161.8
Carbon monoxide (mg/Ncm [24-h] or [97% of 10-min values])	134	134	127.1	21.4	40.5	81.5	40.1
Cadmium (µg/Ncm)	30.6	15.3	0.34	0.34	0.11	0.92	1.07
Lead (µg/Ncm)	336	107	7.1	<1.91	13.6	31.1	26.8
Mercury (µg/Ncm or% removal)	21.4	38.2	1.99	4.59	0.99	1.8	3.68
	85%	85%	98.4%	97.0%	98.7%	97.8%	97.8%
Dioxins/furans (ng/Ncm [NATOTEQ] ng/Ncm [tetra-octa])	No standard	No standard	0.115	0.062	0.026	Not recorded	0.0465 EPA TEF
	22.9	9.9	7.57	5.89	2.75	3.9	2.52

The combustible content of raw MSW is about 70%. Typical noncombustibles are moisture, which may be 10% or more and ash, typically 20%. The moisture and ash-free heating value is about 9,000 Btu/lb, which can be compared with coal at about 12,000 Btu/lb, and oil having about 19,000 Btu/lb.

Table 4 shows typical constituents of MSW and their individual contributions to heating value. Multiplied by the mass percentage, the contribution of each constituent to the total heating value of the waste is obtained. The total is found to be 5,019 Btu/lb, which is fairly typical of contemporary waste, before recycling. The moisture-and-ash-free heating value is found to be 9,069 Btu/lb, which may be compared with coal, as received, at 9,000–12,000 Btu/lb. Recycling removes glass and metals, which do not contribute energy. Removing paper and plastics by recycling, to the extent practical, may actually result in a product with relatively higher heating value as fed to the boiler.

Analysis

The properties of MSW vary significantly, unlike those of fossil fuels. The properties of RDF are less variable due to processing and the inherent reduction in moisture. Table 5 shows the degree of variation of average daily values of RDF measured as it was sent from the storage bin to the boilers at the RDF plant owned by the City of Ames. Volatiles and fixed carbon amount to about 50%, with 23% moisture and 17.4% ash. The ultimate analysis was 30.9% carbon, 4.8% hydrogen, and 22.9% oxygen. The coefficient of variation, the percentage of the standard deviation to the mean, gives the best indication of property variation. Ash fusion temperatures, expressed in degrees Fahrenheit, give a useful indication of the temperatures at which the ash fuses, hence tends to form slag on the surfaces of the boilers. The Ames data indicates that fusion takes place at temperatures higher than about 1,100–1,200°F. Present-day furnaces are generally designed to cool the combustion gases to 1,000–1,400°F before the gases pass through the superheater tube passages where slagging is most likely to be a problem.

Table 6 gives a comparison of typical “reference” values of the composition and the characteristics of MSW and RDF, based on data obtained from operation of the Palm Beach RDF facility. Processing increases the

heating value from 4,720 Btu/lb for MSW to 5,500 Btu/lb for RDF. Also important is the “yield” or fuel energy recovery, which is shown to be 83% for the RDF and 96% for the MSW.

Determining the Heating Value of MSW or RDF

The heating value of the MSW or RDF to be used to generate heat or power is essential to the calculation of the design and performance of the plant [1]. The heating value (HV) may be determined by (Table 7):

- **Boiler as calorimeter:** The boiler must be fully instrumented so that the heat liberated by the combustion, measured by steam generation, and the weights of feed and ash residues can be measured accurately, taking into account moisture added and lost. The ASME Boiler Code may be used for this determination.
- **Laboratory calorimeter:** The standard laboratory oxygen-bomb calorimeter can accept 1 g of milled refuse. Larger bomb calorimeters accept 2.5 g. The National Bureau of Standards (NBS) has developed a large flow calorimeter that can take 2.5 kg samples, permit heating value determinations of coarsely shredded MSW or RDF (see ASTM E-38 on Resource Recovery).
- **Calculate heating value from ultimate chemical analysis.** The weight component of each chemical component is multiplied by its heat of reaction and summed to obtain the total heating value.

The *higher heating value* (HHV) is the amount of heat that *would* be recovered by a boiler if the boiler condensed the water vapor in the combustion products. However, this is not generally the case, in reality, so the HHV is fictitious, although it is commonly used in the USA. Thus:

HHV = 14,447 (carbon) + 60,958 (Hydrogen-oxygen/8) + 33,980 (sulfur) + 1,040 (nitrogen), Btu/lb

The oxygen content is divided by 8 and subtracted from the hydrogen in order to take into account the portion of the hydrogen that has already reacted with oxygen and hence does not contribute heating value. It is convenient to rewrite this equation as follows:

HHV (Btu/lb) = 144.5 (%C) + 609.6 (%H) – 76.2 (%O) + 40 (%S) + 10 (%N)

Waste-to-Energy using Refuse-Derived Fuel. Table 4 Conversion of constituents to heating value by data [1, 23]

Conversion of constituents to heating value by data analysis ^a												
Constituent	As-received analysis					Products		Ultimate analysis				
	Wt%	% MC	% ASH	Btu/lb	Wt × Btu	Wt × MC	Wt × Ash	C	H	O	N	S
Newspaper	10.33	6.0	1.4	7,974	824	0.62	14.8	49.1	6.1	43.0	0.05	0.16
Brown paper	6.12	5.8	1.0	7,256	444	0.35	6.2	44.9	6.1	47.8	0	0.11
Magazine	7.48	4.1	22.5	5,254	393	0.31	168.0	32.9	5.0	38.6	0.07	0.09
Corrugated	25.68	5.2	5.1	7,043	1,809	1.34	130.0	43.7	5.7	44.9	0.09	0.21
Plastic coated	0.84	4.7	2.6	7,341	62	0.04	2.2	45.3	6.2	45.5	0.18	0.08
Waxed milk cartons	0.74	3.5	1.2	11,327	95	0.03	1.0	59.2	9.3	30.1	0.12	0.10
Paper cartons	2.27	6.1	6.5	7,258	165	0.14	41.8	44.7	6.1	41.9	0.15	0.16
Junk mail	3.03	4.6	13.1	6,088	184	0.14	40.0	37.9	5.4	42.7	0.17	0.09
Tissue paper	2.18	7.0	0.9	7,000	153	0.15	19.6	43.6	6.1	49.0	0.16	0.14
Food wastes	2.52	78.3	1.1	1,795	45	1.97	2.7	49.0	6.6	37.6	1.7	0.20
Citrus rinds	1.68	78.7	0.7	1,707	29	1.32	1.2	48.0	5.7	41.7	1.1	0.12
Cooked meats	2.52	38.7	3.1	7,623	192	0.97	7.8	59.6	9.5	24.7	1.0	0.19
Fried fats	2.52			16,466	415			73.1	12.0	14.3	0.4	0.07
Leather	0.42	7.5	21.2	7,243	30	0.03	8.9	42.0	5.3	22.8	6.0	1.0
Rubber	0.42	1.2	29.7	10,900	46	0.01	12.5	53.2	7.1	7.8	0.5	1.3
Plastics	0.84			15,910	134			78.0	9.0	13.0		
% Combustible 69.60				HHV = 5019		↓	↓	41.7	5.8	27.6	2.8	0.25
% Moisture in combustibles:						7.42		↓	↓	↓	↓	↓
% Ash in combustibles:							13.2					
HHV by ultimate analysis:	Ultimate analysis	= 79.4% =						41.7	5.8	27.6	2.8	0.25
	HHV of elements	=						145	609	-76	40	10
	HHV × Wt.%	= 7,595.5 =						6,046	3,532	2,097	112	2.5
	HHV3(MAF)	= 7,596/(1-0.132)					= 8,751 Btu/lb					
HHV by components:	HHV1 (as received)	= 5,019/(0.696)					= 7,201 Btu/lb					
	HHV2 (dry)	= 7,201/(1-0.074)					= 7,776 Btu/lb					
	HHV3 (MAF)	= 7,201/(1-0.074-0.132)					= 9,069 Btu/lb					

Waste-to-Energy using Refuse-Derived Fuel. Table 5 Variability of daily values of characteristics of RDF - Ames Iowa [8]

Characteristic (%)	$\bar{X}$	Variability (about mean at 95% confidence coefficient) ($\pm$)	S_x standard deviation	Coefficient of variation (%)
Heating value (kv/kg)	13.050	590	1021.6	7.8
Bulk density (kg/m ³)	157	10.5	18.1	14.1
Proximate analysis (%)				
Moisture	23.0	2.4	4.2	18.3
Ash	17.4	1.8	3.2	18.3
Volatile	54.6	2.1	3.7	6.8
Fixed carbon	4.95	1.4	2.4	48.6
Ultimate analysis (%)				
Carbon	30.9	1.3	2.2	7.2
Hydrogen	4.8	0.2	0.3	6.8
Oxygen	22.9	2.3	3.9	17.1
Sulfur	0.43	0.1	0.2	44.8
Chlorine	0.24	0.06	0.11	45.1
Nitrogen	0.42	0.06	0.11	25.4
Ash analysis (%)				
SiO ₂	48.19	2.3	4.1	8.4
Al ₂ O ₃	11.75	1.3	2.3	19.5
Fe ₂ O ₃	4.3	7.7	1.3	31.0
TiO ₂	1.45	0.15	0.26	17.74
P ₂ O ₅	0.79	0.16	0.28	34.9
CaO	12.71	0.93	1.7	12.7
MgO	2.38	0.18	0.31	13.1
Na ₂ O	4.37	0.35	0.60	13.7
K ₂ O	1.8	0.14	0.24	13.6
Ash fusion (°C)				
<i>Reducing atmosphere</i>				
Initial deform	1,106	72	42	3.8
Softening	1,143	31	18	1.6
Hemispherical	1,158	41	24	2.0
Fluidization	1,177	65	37	3.2
<i>Oxidizing atmosphere</i>				
Initial deform	1,149	41	23	2.0
Softening	1,165	50	29	2.5
Hemispherical	1,181	64	37	3.1
Fluidization	1,197	79	46	3.8

Waste-to-Energy using Refuse-Derived Fuel. Table 5 (Continued)

Characteristic (%)	$\bar{X}$	Variability (about mean at 95% confidence coefficient) ($\pm$)	S_x standard deviation	Coefficient of variation (%)
Particle size				
Geometric mean (mm)	12	1.2	1.4	13.0

Waste-to-Energy using Refuse-Derived Fuel. Table 6

Typical Reference values of Refuse [13]

	Reference	
	MSW (% by wt)	RDF (% by wt)
<i>Component analysis</i>		
Corrugated board	5.53	–
Newspaper	17.39	–
Magazines	3.49	–
Other paper	19.72	–
Plastics	7.34	–
Rubber, leather	1.97	–
Wood	0.84	–
Textiles	3.11	–
Yard waste	1.12	–
Food waste	3.76	–
Mixed combustibles	17.75	–
Ferrous	5.50	–
Aluminum	0.50	–
Other nonferrous	0.32	–
Glass	11.66	–
Total	100.00	
<i>Ultimate analysis</i>		
Carbon	26.65	31.00
Hydrogen	3.61	4.17
Sulfur	0.17 (Max. 0.30)	0.19 (Max. 0.36)
Nitrogen	0.46	0.49
Oxygen	19.61	22.72
Chlorine	0.55 (Max. 1.00)	0.66 (Max. 1.20)
Water	25.30	27.14

Waste-to-Energy using Refuse-Derived Fuel. Table 6 (Continued)

	Reference	
	MSW (% by wt)	RDF (% by wt)
Ash	23.65	13.63
Total	100.00	100.00
Heating value	4,720 Btu/lb (10,979 kJ/kg)	5,500 Btu/lb (12,793 kJ/kg)
Fuel value recovery (% MSW)	96	
Mass yield (% RDF/MSW)	83	

$$\text{HHV (MJ/kg)} = 0.336 \text{ C} + 1.418 \text{ H} - 0.148 \text{ O} + 0.094 \text{ S} + 0.059 \text{ N}$$

As-Received, Dry and Moisture and Ash-Free Heating Values

- As-received heating value:
 $\text{HHV1} = \text{HHV3} \times (1 - (\text{ash fraction}) - (\text{moisture fraction}))$
- Dry heating value: $\text{HHV2} = \text{HHV3} \times (1 - (\text{ash fraction}))$
- Moisture and ash-free heating value:
 $\text{HHV3} = (\text{HHV1}/(1 - \text{AF-MF})) = \text{HHV2}/(1 - \text{AF})$

Air Required for Combustion

The air required for complete combustion of the combustible components of MSW or RDF can be calculated from the chemical equations, heat release and atomic weights of the components, expressed in pounds of air per million Btu of heat released by dividing by the HHV, as follows:

$$\text{Theoretical air (TA)} = \{138.17 (\text{C}/12 + \text{H}/4 - 0.32 + \text{S}/32 + \text{N}/28) \times 10^6\} / \text{HHV, lb/million Btu}$$

Waste-to-Energy using Refuse-Derived Fuel. Table 7 Calculation of air required for combustion of RDF [1]

Component	% Wt	×	Heat release	=	HHV	% Wt	×	Factor	AIR
Carbon	26	×	144.47	=	3,756.22	26	×	11.51	299.26
Hydrogen	3.8	×	609.58	=	2,316.40	3.8	×	34.5	131.1
Oxygen	21.21	×	−76.2	=	−	21.21	×	−4.31	−91.63
Sulfur	0.18	×	39.82	=	7.17	0.18	×	4.32	9.78
Nitrogen	0.53	×	10.5	=	5.57				
Chlorine	0.18								
Nitrogen	0.4								
Moisture	26.54								
Ash	21.7								
Totals:	100	HHV = 4,469.15 Btu/lb							348.51

This form may be more convenient to use:

Theoretical air (TA) = $\{[11.51 (\%C) + 34.5 (\%H) - 4.32 (\%O) + 4.3 (\%S)] \times 10^6\} / \text{HHV}$

It is convenient to use a tabular form to make these calculations, as follows, using typical values for weight percentages.

The value of theoretical air per unit of heat released by combustion, determined in this manner from chemical analysis of samples of MSW and RDF, appears to be independent of the nature of the waste, hence may be used with confidence in estimates of combustion air requirements.

Theoretical air (TA) = $\{348.5 \times 10^4\} / 4,469.15 = 760$ lb air per million Btu

Lower Heating Value. While the HHV is commonly used in the US, in other countries the lower heating value (LHV) is used. It can be calculated from the higher heating value (HHV) by subtracting the latent heat of evaporation of the water in the products of combustion. This water includes the water content of the fuel and the water produced by the combustion of hydrogen:

$\text{LHV} = 144.5 (\%C) + 609 (\%H) - 76 (\%O) + 40 (\%S) + 10 (\%N) - 10.5 (\%H_2O)$

The water produced by the combustion of hydrogen is nine times the weight of hydrogen in the fuel $[(16 + 2)/2 = 9]$. The hydrogen in the fuel is known from the ultimate analysis. For refuse fuels, the

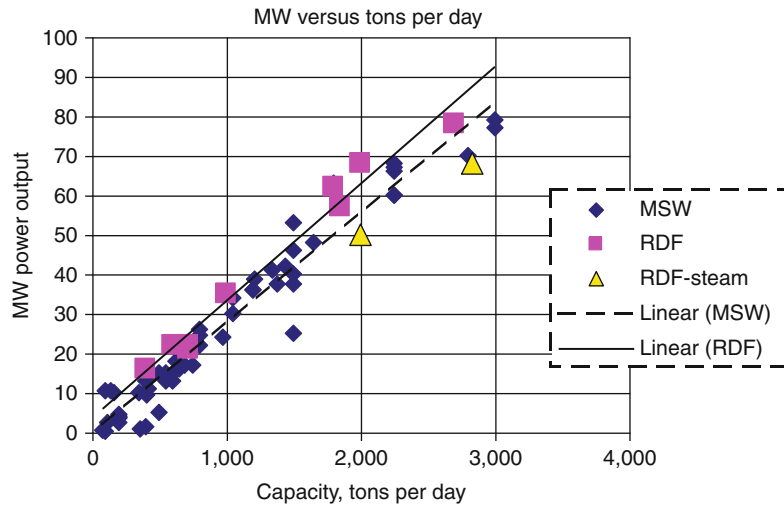
hydrogen is about 7.5% of the combustible matter. The latent heat at atmospheric pressure is 1,040 Btu/lb.

Loss of heating value due to not condensing water vapor = $7.5 \times 9 \times 10.50 = 712$ Btu/lb combustible.

Power Production from Refuse Fuel

The energy yield of the Processed Refuse Fuel (PRF) process employed at the SEMASS Resource Recovery Facility would be expected to be higher, closer to the mass-burn yield. The simpler SEMASS process, with minimal processing, has only one stage of shredding and therefore uses less electric power and is simpler to operate and maintain. The SEMASS process has a higher overall plant efficiency than one burning more highly processed refuse, since some combustible material is lost due to the additional processing. RDF plants may be expected to have a higher thermal efficiency since the moisture content of the fuel is greatly reduced below that as received from the refuse trucks. The power used for shredding directly reduces the moisture content, hence is not wasted.

Other factors, such as the amount of excess air required and the steam pressures and temperatures at which the boiler generates the steam, make it difficult to compare the two approaches to processing municipal waste. The kilowatt-hours produced per ton of waste taken into the plant is the most telling ratio



Waste-to-Energy using Refuse-Derived Fuel. Figure 33
Power produced per ton of waste (MW/TPD)

used in evaluating overall performance. This ratio varies throughout the year, since it depends on the heating value of the waste. A comparison between RDF and MSW facilities has been made using the design rating, nominal tons per day, at standard reference values of heating value, listed in Table 6. This is shown graphically in Fig. 33.

Future Directions

The technology for processing MSW and commercial waste to recover materials and generate energy has evolved since the 1970s and, 40 years later, is positioned to help achieve national and international goals of greenhouse-gas reduction and renewable energy generation. Concerns related to off-shore drilling, coal mining, foreign oil sources, nuclear waste disposal, and environmental contamination from landfills have forced a rethinking of waste as a viable fuel source.

One ton of fuel created from waste can produce the same amount of energy as 1 t of coal or a barrel of oil. “Replacing just 10% of the coal used to generate electricity in the United States in 2007 with RDF would create a potential demand for 225 million tons of RDF – significantly more than might be available if all the waste currently destined for landfills were diverted for RDF processing” [23].

Bibliography

Primary Literature

1. Hasselriis F (1984) Refuse-derived fuel processing. Butterworth Publishers, Boston (available from author)
2. Sutin GL (2004) The evolution of processed refuse fuel technology. Energy Answers Corporation, Albany, New York
3. Sutin GL (1998) Suspension firing – the heart of an integrated waste-to-energy system. Energy Answers Corporation, Albany
4. Rochford RS, Witkowski WJ (1978) Considerations in the design of shredded municipal refuse burning and heat recovery system. In: Proceedings of the 8th National Waste Processing Conference, Chicago, ASME, New York
5. France HA (1986) Hamilton SWARU retrofit. In: 1986 National Waste Processing Conference, ASME, Denver
6. Fiscus DE et al (1977) St. Louis demonstration final report. EPA-600/2-77-155a, MERL/ORD/U.S. EPA, Cincinnati
7. Gorman PG et al (1977) St. Louis Demonstration final report: power plant equipment, facilities, and environmental evaluations. U.S. Environmental Protection Agency Report EPA-600/2-77-1556
8. Hasselriis F (1983) Thermal systems for conversion of municipal waste, Vol. 4, Burning refuse-derived fuel in boilers. A Technology Status Report. ANL/CNSV-TM-120, Argonne National Laboratory
9. Fiscus DE et al (1983) Study of existing RDF-cofiring experience. Appendixes to Phase I Final Report, ANL/CNSV-TM-134, vol 2, Argonne National Laboratory

10. Hasselriis F (1980) The greater Bridgeport, Connecticut waste-to-power system. In: Proceedings of the 1980 National Waste Processing Conference, Paper Number WSS/CI 81-6
11. Energy Recovery Council. <http://www.energyrecoverycouncil.org>
12. Michaels T (2007) The 2007 IWSA directory of waste-to-energy plants: Integrated Waste Services Association, 31 p. http://www.wte.org/docs/IWSA_2007_Directory.pdf
13. Di Maria F, Pavesi G (2006) RDF to energy plant. . . energetic and economical analysis. *Appl Therm Eng* 26:1291–1300
14. Murphy ML (1999) Design and performance requirements for a fluidized bed boiler firing municipal refuse derived fuel in Ravenna, Italy. In: Proceedings of the 15th International Fluidized Bed Combustion Conference, Savannah, 16–19 May 1999, Paper FBC99-0128, epi2@energyproducts.com
15. Pope K (1998) Difficult fuels and stringent emissions requirements influence combustion technology preference in Europe, Energy Products of Idaho, epi2@energyproducts.com
16. Abrams RF et al (2010) 2,400 tons per day refuse-derived fuel facility with advanced boiler and air pollution control systems. In: Proceedings of the 18th Annual North American Waste-to-Energy Conference, Orlando, Florida
17. Abrams RF, Faia R (2009) RSCR system to reduce NO_x emissions from boilers. In: Proceedings of the 17th North American Waste-to-Energy Conference, Chantilly, Virginia
18. Ferraro FA et al (1988) Results of emissions and ash testing at the maine energy recovery WTE plant. WTI Energy, Inc., Biddeford, Maine
19. Bresowar GE et al (1988) Operating experience with a spray dry scrubbing system on a RDF-Fired boiler. Combustion Engineering, Inc.
20. Entropy Environmentalists Inc. (1988) Bridgeport Resco, Co., LP resource recovery facility compliance test report. Wheelabrator Environmental Systems, Inc.
21. J. L. Hahn, H. P. Von Dem Fange, D. Sofaer, Ogden Projects, Inc., "Recent Air Emissions Data from Ogden Martin Systems, Inc. Resource Recovery Facilities, Which Became Operational During 1988," June 1989, covantaenergy.com
22. Kaiser E (1966) Chemical analysis of refuse components. In: Proceedings of the 1966 National Incinerator Conference, American Society of Mechanical Engineers, New York
23. Gershman HW (2010) Fuel for the fire: a renewable energy push could spark demand for refuse-derived-fuel. *Waste Age*, March 2010
24. Mahoney PF, Zakaria J (1993) Basic design and operation information of the semass resource recovery facility. In: RDF Technology Seminar, Williamsburg, 29 Mar 1993
- Bobman M, Culbertson J (2010) Waste composition in the Northeast U.S. – implications for resource recovery. In: Proceedings of the 18th Annual North American Waste-to-Energy Conference, NAWTEC, Orlando
- Consonni S, Giuliano M, Grosso M (2005) Alternate strategies for energy recovery from municipal solid waste. In: WTER 2005, Columbia University
- Eco-Industrial Parks as a Vehicle for Converting Liability to Productive Redevelopment. 18th AHC Group Corporate Affiliates Workshop, Saratoga Springs, NY, June 2008
- Efficient Recovery of Materials and Energy from Residual Waste. R'07 World Congress, Davos, Switzerland, September 2007
- Energy Answers International (2010) Fairfield renewable energy plant and resource recovery project. *Energy Answers International*, EnergyAnswersInternational.com
- Gershman HW (2010) Fuel for the fire: a renewable energy push could spark demand for refuse-derived-fuel. *Waste Age*, March 2010
- Hasselriis F (1999) Practical design of waste incineration (Chapter 2.3) and Calculations for permitting and compliance (Chapter 2.4). In: Lee CC (ed) McGraw-Hill handbook of environmental calculations, ISBN 0-07-038183-6
- Hollander H, Kieffer JK, Eller V, Stephenson JW (1980) A comprehensive municipal waste characterization program. In: Proceedings of the 1980 National Waste Processing Conference, ASME, New York
- Lisiecki HG (1978) RDF storage and retrieval problems, cause-effect options. In: Proceedings of the National Waste Processing Conference, Chicago, ASME, New York
- Mahoney PF (1980) ANSWERS: a comprehensive solution to the economic, technical and institutional problems of solid waste reuse. In: 8th Annual Industrial Pollution Conference, Houston
- Mahoney PF, Mullen J (1988) Approaching total recovery in municipal solid waste to energy systems. In: 23rd Inter-Society Energy Conversion Engineering Conference, Denver
- Murphy ML (2002) Advancing waste to energy technology design and performance of EPI fluidized bed RDF-fired power plants worldwide. In: 10th North American Waste to Energy Conference, ASME, Philadelphia, NAWTEC10-1011, epi2@energy-products.com.
- Nollet AR, Greeley RH (1982) New concepts for explosion alleviation in shred-first solid waste plants. In: Proceedings of the National Waste Processing Conference, ASME, New York
- Providing Market Competitive Sustainable Solutions Drive the Bottom Line. AHC Corporate Affiliates Benchmarking Workshop, Phoenix, AZ, February 2007
- Resource Recovery vs. Waste to Energy. Confederation of European Waste-to-Energy Plants Congress, Vienna, Austria, May 2006
- Solid Waste as an Economic Generator for Sustainable Development. AHC Group 9th Annual Affiliate Program, Saratoga Springs, NY, June 2001
- Tchobanoglous G, Kreith F (eds) (2002) Handbook of solid waste management. McGraw-Hill, New York

Books and Reviews

Beyond Sustainable: The Efficient Extraction of Energy & Marketable Products from Waste Materials. Keynote address – Waste Management & Recycling Congress, Berlin, Germany, October 2008

Waste-to-Energy, Introduction

NICKOLAS J. THEMELIS

Earth Engineering Center, Columbia University,
New York, NY, USA

Article Outline

Municipal Solid Wastes (MSW)
Composition and Heating Value of MSW
Chlorine Concentration in MSW
Waste Reduction
Collection and Processing or Disposal of MSW
Recycling
Composting
Dealing with Mixed MSW and Post-recycling Residues
Energy Recovery from Wastes (Waste-to-Energy
or WTE)
Landfilling
The Hierarchy of Waste Management
Greenhouse Gas Effects of Waste Management
Thermal Treatment Technologies
Grate Combustion of “As-Received” MSW
Combustion of Refuse-Derived Fuel (RDF)
Emerging WTE Technologies
Beneficial Use of WTE Ash
Need for Government Support
Bibliography

When I was asked by the Editor of the Encyclopedia of Sustainability Science and Technology, to put together a section on the role of waste-to-energy, I agreed to do it because the subject of resource recovery from the discards of society is an essential part of what the United Nations has defined as “sustainable development.” Prior to the second half of the twentieth century, the consumption of materials and energy by humanity was a small fraction of what it is today. An example of the rapid growth in consumption is the use of copper, a metal that has been known since the beginning of history: A 1995 study by the author showed that more copper was consumed between 1945 and 1995 than in the preceding six millennia (Fig. 1).

Technology has made possible this rapid economic development and, also, the accompanying degradation of the global environment and the loss of innumerable

species. However, in many cases technology has also helped immeasurably in mitigating or slowing down the adverse effects of economic development. On the basis of my studies and those of my graduate students in the past 15 years, I am convinced that waste-to-energy technology is definitely one of those bright spots where technology and ecology are not adversaries.

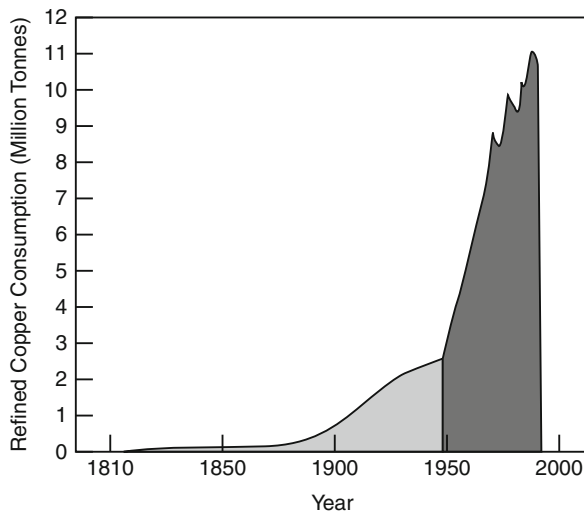
As noted in the excellent essay by William Ophuls “Plato’s Revenge” (MIT Press 2011), there are similarities between the present age and that of the conquistadores: Scientists and engineers invent ways for producing the goods and arms that merchants and soldiers use to dominate markets and nations.

A very clear case of what can be called ecological imperialism is the massive transformation of greenfields to landfills. In the past, landfills were located in, or near, urban centers so there was a good reason for cities to manage their wastes and try to conserve land. In recent times, some cities, such as New York, have run out of landfill space and transport their wastes hundreds of kilometers away to mega-landfills. This is made possible by waste transfer stations, diesel trucks, and other technological equipment and, also, by the fact that cities are generally more affluent than the communities where the mega-landfills are created.

Fortunately, in the last 50 years, engineers have developed an alternative to landfills: waste-to-energy power plants that can be built within or near cities and transform the non-recycled residues of society to electricity, heat, and metals. Closing the materials loop in this way is a good example of industrial ecology using technology that mimics the age-long ecological cycle.

In developing the waste-to-energy (WTE) section of this Encyclopedia, the editor invited specified contributions from some of the most knowledgeable and prominent scientists and engineers in the area of integrated waste management.

- The essay by Johannes Martin and Ralf Koralewska of Martin GmbH (Germany) (► [Martin Waste-to-Energy Technology](#)) describes one of the oldest and constantly developing WTE technologies that is used in hundreds of plants in Europe, Asia, and North America.



Waste-to-Energy, Introduction. Figure 1
Copper consumption in the twentieth century

- Craig Kedrowski and Peter Chromec of AE&E Inc. (USA) discuss the VonRoll grate combustion technology that is also used extensively throughout the world (► [Hitachi Zosen Inova Technology](#)). Both of the above technologies are based on the combustion of as-received municipal solid wastes (MSW) on a moving grate.
- Pat Mahoney of International EnergyAnswers Corp. (USA) and Floyd Hasselriis of Hasselriis Associates (USA) describe technologies that process the MSW mechanically before it is fed to the combustion chamber (► [Waste-to-Energy using Refuse-Derived Fuel](#)). Such technologies are called refuse-derived fuel (RDF) processes.
- The third largest waste-to-energy technology is fluidized bed combustion, discussed by Franz P. Neubacher (► [Waste-to-Energy: Fluidized Bed Technology](#)), author of the book *Waste-to-Energy in Austria*.
- Processes developed in Japan for converting MSW to gaseous or liquid fuels are the subject of the essay by Kunio Yoshikawa of the Tokyo Institute of Technology (► [Gasification and Liquefaction Alternatives to Incineration in Japan](#)).
- The description of various WTE technologies is completed by the essay of Nickolas Themelis of Columbia University and Armelle Vardelle of the University of Limoges that discusses thermal plasma-assisted WTE processes (► [Plasma-Assisted Waste-to-Energy Processes](#)).
- The transport and reaction phenomena in the combustion chamber are dealt with by J. Swithenbank and V.N. Sharifi of Sheffield University (UK), the most prominent experimental group on WTE (► [Incinerator Grate Combustion Phenomena](#)).
- The very important role of WTE facilities as power generators is treated by M.A.J. (Marcel) van Berlo of the Free University of Amsterdam and co-developer of the highly efficient WTE plant of AEB Amsterdam (NL) (► [Waste-to-Energy Facilities as Power Plants](#)).
- Increasing the thermal efficiency of WTE facilities by using the “waste” stream for district heating (► [Waste-to-Energy for District Heating](#)), as practiced extensively in Denmark and some other nations in northern Europe, is discussed by Lasse Tobiasen and Bettina Kamuk of Ramboll Energy (Denmark).
- Dieter O. Reimann of the Confederation of European WTE plants reviews the material and energy balances in WTE facilities and the energy-production performance of European waste-to-energy plants (► [Waste-to-Energy: Energy Resource in Solid Wastes](#)).
- The beneficial use of WTE ash in Europe (WTE Ash Management in Europe) is discussed by Jurgen Vehlow of Karlsruhe Research Center (Germany) and in North America (WTE Ash Management in United States) by Floyd Hasselriis of Hasselriis Associates.
- Bernd Bilitewski and Christoph Wunsch of the Technical University of Dresden (Germany) describe the essential role of WTE in reducing greenhouse gas emissions (GHG) (► [Greenhouse Gas Emission Reduction by Waste-to-Energy](#)).
- The essay of Brian Babor and Michael Van Brunt of Covanta Energy (USA) examines how integrated waste management can be part of the solution to the emerging crises of climate change and inadequate supply of renewable energy for future generations (► [Waste Management for Sustainable Society](#)).
- An LCA comparison of the environmental impacts of WTE to sanitary landfilling (LFGTE)

is presented in the essay by P. Ozge Kaplan, Joseph F. DeCarolis, and Morton A. Barlaz of North Carolina State University (USA), the foremost US research group on sanitary landfilling (► [Life Cycle Comparison of Waste-to-Energy to Sanitary Landfill](#)).

- Helmut Rechberger of Vienna University of Technology (Austria) shows how WTE decreases the entropy of solid wastes and increases metals recycling (► [Waste-to-Energy: Decreasing the Entropy of Solid Wastes and Increasing Metal Recovery](#)). Paul H. Brunner, also of Vienna University of Technology and winner of Columbia's 2006 WTERT Award, completes this extraordinary collection of essays, by explaining why thermal treatment of solid wastes is a key element of sustainable waste management (► [Thermal Treatment of Waste: Key Element for Sustainable Waste Management](#)).

I am grateful to all the authors for their contributions to this section. In the rest of this introductory entry, I will take the opportunity to summarize my own understanding of “sustainable waste management” and waste-to-energy, as derived from several dozens of studies and graduate theses sponsored by the Earth Engineering Center of Columbia University (EEC; www.eecny.org) and the Waste to Energy Research and Technology Council (WTERT; www.wtert.org), both headquartered at Columbia University.

Winner of Columbia/WTERT 2006 Industry Award: ASM Brescia, Italy



Municipal Solid Wastes (MSW)

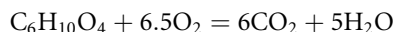
Most of the solid wastes can be classified into municipal (residential and commercial), industrial, construction and demolition, and agricultural wastes. Municipal

solid wastes (MSW) are the most heterogeneous since they consist of the residues of everything used by humanity: food and plant wastes, used furniture, papers, plastics, fabrics, leather, metals, glass, and other materials.

Composition and Heating Value of MSW

The composition of MSW varies somewhat from urban to rural settings and also from country to country. Table 1 shows a typical composition of MSW in a US city. The largest constituent is paper; other low-moisture organic materials are plastics, textiles, rubber, leather, and wood. These materials can be called “dry combustibles,” in distinction to the “wet combustibles” of food, plants, and other wastes that may contain up to 70% water. The “noncombustible” wastes are metal, glass, and other inorganic compounds. Large items such as appliances and furniture can be broken down to recyclable metal or combustible wood and plastics. Hazardous wastes, such as paints, oils, and chemicals constitute only 0.4% of the total waste stream and are handled separately.

On the basis of the chemical composition of each material in Table 1 and the atomic weights of their chemical elements, Themelis et al. (2002) [1] showed that the chemical formula $C_6H_{10}O_4$ most closely approximated the mix of organic compounds in MSW. This formula corresponds to that of about ten organic compounds, such as adipic acid and ethylene glycol diacetate, whose heat of formation is about -960 kJ/mol (Roinen 1999). Using this value, the combustion of the $C_6H_{10}O_4$ “compound” can be represented by the chemical equation



The above reaction is highly exothermic and at the reaction temperature of $1,000^\circ\text{C}$ generates about 2.7 MJ/mol. Since the gram mol weight of $C_6H_{10}O_4$ is 0.146 kg, the “theoretical” heat of oxidation of the combustible components of MSW, that is, excluding moisture and inorganic materials, is calculated to be 18.5 MJ/kg. In the actual case where moisture and inorganic compounds are present, the average heat content of MSW in E.U. is about 10 MJ/kg. In terms

Waste-to-Energy, Introduction. Table 1 Composition of New York City MSW

Waste component	% weight
Paper (31.3%)	
Corrugated cardboard	4.7
Newspapers	9.2
All other paper	17.4
Plastics (8.8%)	
HDPE (clear and colour)	1.1
Films and bags	4.8
PET	0.5
Polypropylene, polystyrene	0.9
PVC	0.1
All other plastics	1.4
Wood	2.2
Textiles	4.7
Rubber and leather	0.2
Fines	2.3
Other combustibles	2.3
Food waste	12.7
Grass/leaves	3.4
Brush/prunings/stumps	0.7
Disposable diapers	3.4
Miscellaneous organics	7.8
Glass	5.0
Clear glass containers	2.9
All other glass	2.1
Aluminium	0.9
Ferrous metal	3.9
Hazardous waste	0.4
Bulk Items (appliances, furniture, etc)	5.0

of power generation potential, this value corresponds to 2.86 MWh of chemical energy per ton of typical MSW.

The effect of various waste materials and moisture on the heating value of MSW is illustrated in Fig. 2.

Chlorine Concentration in MSW

A troublesome element in MSW is chlorine because of its corrosive effect on boiler tubes. Table 2 shows that MSW contains much more chlorine than coal. This obliges WTE power plants to operate at lower steam temperatures and thus thermal efficiencies than coal fired plants.

The contribution of the various components of a typical US MSW to its chlorine concentration is shown in Table 3.

Waste Reduction

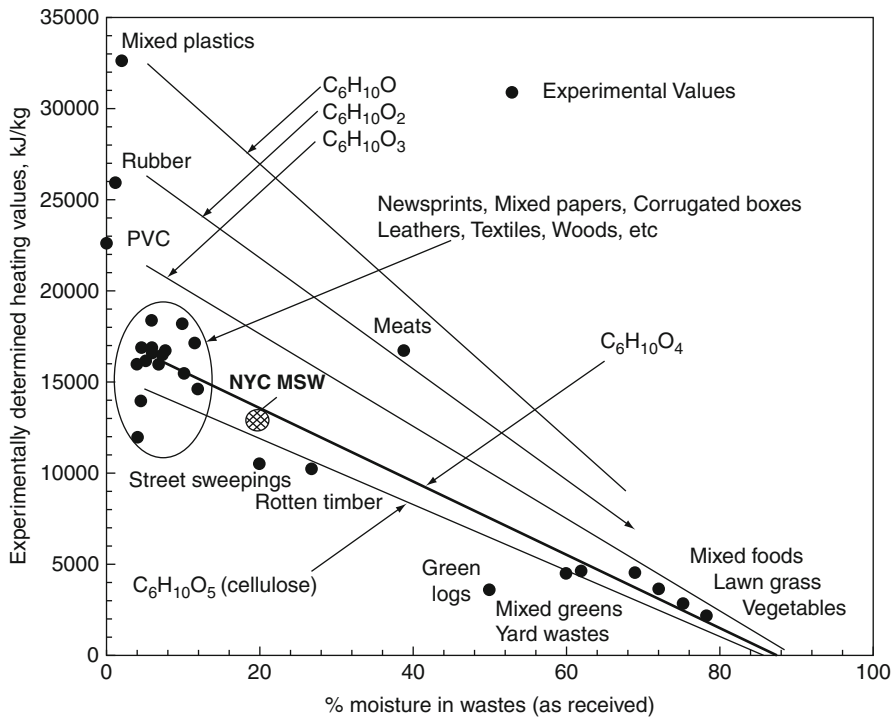
Economic development and prosperity are accompanied by the generation of more solid wastes per capita. The generation of wastes can be reduced to some extent by improved design of products and packaging materials, for example, by increasing intensity of service per unit mass of material used. However, the major factor is culture and lifestyle, as indicated by the fact that US citizens generate twice as much solid waste per capita, as northern Europeans and Japanese who live just as well. However, even in material-thrifty countries, there remains a large amount of solid wastes to be taken care of.

Collection and Processing or Disposal of MSW

Sustainable management of municipal solid wastes (MSW) requires that every possible effort be made to separate recyclable or compostable materials from the MSW stream. Experience has shown that it is best for these materials to be separated at the source, that is, at households, businesses, and institutions. The cost of source separation is then shared by the citizens (in terms of time and effort to sort out the recyclables) and the municipality (in terms of separate collection vehicles and processing systems).

Recycling

A 2009 study by Columbia University for Recycle America (USA) showed that it is environmentally and economically better to have a “single stream” rather than “multiple stream” collection of paper, metals, and designated plastic and glass materials. The single stream of recyclables is then collected by the municipality or a private company and brought to a Materials



Waste-to-Energy, Introduction. Figure 2
Effect of constituents and moisture on calorific value of MSW [1]

Waste-to-Energy, Introduction. Table 2 Chlorine concentration in MSW and in coal

	New York City MSW	US typical MSW	UK typical MSW	US coals, range
Total Cl, g/kg	4.71	7.26	4.53	0.4–1
%	0.471	0.726	0.453	0.04–0.1

Recovery Facility (MRF) where it is separated to marketable streams, by means of electromechanical and manual methods.

It should be kept in mind that, unless the source-separated materials are usable, they will end up in landfills. An example of the lack of markets for certain materials is the fact that less than 10% of the plastic wastes generated in the USA are actually recycled; the corresponding figure in Germany is 20%.

The most easily recyclable materials are ferrous and nonferrous metals. The highest environmental and also economic benefit, per kilogram of material recycled, is derived from the recovery of aluminum because of the large amount of energy required to produce aluminum

from bauxite ore. The largest component in MSW is mixed paper that can be re-pulped to produce wrapping paper and cardboard.

Composting

A second “stream” collected by many municipalities consists of yard or “green” wastes such as leaves, grass, tree branches, etc. Usually, these materials are brought by citizens to a collection center and are composted aerobically (i.e., in the presence of atmospheric oxygen) to carbon dioxide and a compost product that can be used as soil conditioner. Yard wastes can also be shredded and used as Alternative Daily Cover (ADC) in

Waste-to-Energy, Introduction. Table 3 Contribution of MSW components to its chlorine content

Material in MSW	Contribution to chlorine content (%)
Paper	10.1
Plastics	44.7
Organics	17.9
Textiles	9.4
Wood	5.8
Miscellaneous combustible	10.7
Glass	0.4
Metals	0.5
Miscellaneous noncombustible	0.5
Total	100.0

sanitary landfills, in place of soil. For example, nearly two million tons of yard wastes are used as ADC in California and the state considers this use to be equivalent to recycling. On the average, MSW contains 10% of yard wastes and the same or higher percentage of food wastes.

Food wastes generate undesirable odors in open windrow composting but can be composted *anaerobically* (i.e., in the absence of atmospheric air) in digesters to produce a biogas that contains about 50% of methane and, after a curing period, a compost product. However, composting of food wastes has not been successful on a large scale for two main reasons: First, it is not easy for households to collect odorous food wastes that decompose fairly rapidly in storage. Second, food wastes composted aerobically in windrows emit odors; therefore, they must be composted in closed composting systems or digested anaerobically (AD), both of which are costly to build and operate. As a result, in the USA about 50% of the yard wastes are collected and composted, while less than 5% of the generated food wastes are processed.

It should be noted that composting – both aerobic and anaerobic – is practical only for source separated organics. Otherwise, most of the compost product is not marketable as a soil conditioner and in many cases ends up in landfills.

Dealing with Mixed MSW and Post-recycling Residues

International experience has shown that there are only two ways to manage mixed MSW and the unavoidable residues from recycling and composting: (a) thermal processing, where the chemical energy of the wastes is recovered in the form of steam or syngas fuel, and (b) modern (sanitary) landfilling where some of the chemical energy in the MSW may be recovered in the form of landfill gas (LFG; 50% methane), thus reducing landfill methane emissions by 50–75%.

Energy Recovery from Wastes (Waste-to-Energy or WTE)

As stated earlier, energy is stored in chemical form in all waste materials that contain organic compounds containing carbon and hydrogen. These compounds are either of biogenic origin (e.g., paper, wood, food and plants, cotton, leather) or are anthropogenic materials created from fossil fuels, such as plastics and other petrochemicals. A carbon isotope analysis of stack gas from several WTE plants by Covanta Energy has shown that about two thirds of the carbon in the US MSW is biogenic and one third anthropogenic. The chemical energy content of metals, glass, and other inorganic materials is negligible.

The principal way of recovering the energy stored in wastes is by controlled combustion on a moving grate. The heat contained in the combustion gases is transferred to water and steam in a boiler and the superheated steam is used in a turbine generator of electricity and/or for district heating.

The chemical energy contained in MSW can also be recovered from wastes by means of partial combustion and gasification. The product of gasification is synthetic gas (“syngas”), consisting of carbon monoxide and hydrogen. It can be used as fuel in a gas engine or turbine generator of electricity or can be converted chemically to a liquid fuel (e.g., by means of the Fischer-Tropsch reaction).

Sorted plastic wastes, which have a high calorific value, can be converted into a liquid fuel by using an external source of heat, such as microwave energy. This process is called pyrolysis.

Landfilling

Landfilling is the ancient way of dealing with solid wastes and is still the principal means of waste management in most nations (Fig. 3). MSW that is not subjected to recycling, composting, or WTE, plus the residues from these processes (e.g., ash from combustion that is not used as construction material), must be disposed in properly designed and operated landfills.



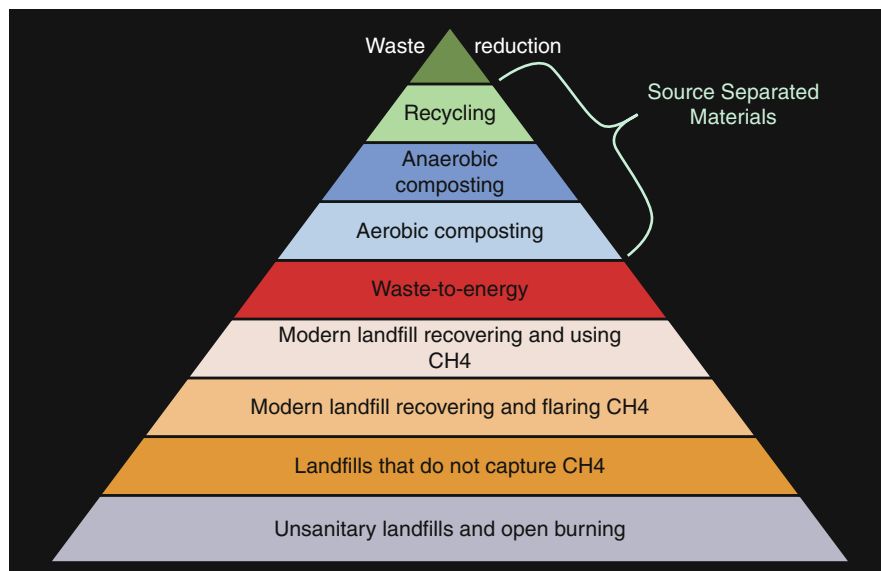
Waste-to-Energy, Introduction. Figure 3
Landfill serving the island of Mauritius

A Columbia University study has shown that the urban MSW landfilled globally amounts to about one billion tons annually. Of this amount, only 200 million tons of MSW are deposited in modern landfills that try to collect some of the biogas generated.

A 2007 study by Columbia University showed that the current growth of WTE capacity, at about four million tons per year, will not be enough to curb landfill methane emissions by the year 2030 because population growth and economic development will result in a much greater rate of MSW generation and landfilling. The only way to reduce landfill greenhouse gases (GHG) between now and 2030 is by achieving a 7.5% growth in thermal treatment capacity on a global scale, or by increasing the amount of methane captured at landfills.

The Hierarchy of Waste Management

The “hierarchy of waste management” is a pyramid that illustrates the priorities that have been established in developed nations with regard to different methods for managing solid wastes. Waste reduction is placed at the top, followed by recycling and composting, waste-to-energy, and landfilling. In 2008, the Earth Engineering Center of Columbia University proposed the expanded hierarchy shown in Fig. 4 that divided



Waste-to-Energy, Introduction. Figure 4
The hierarchy of waste management

composting into anaerobic and aerobic, and landfilling into three categories: Modern landfills, also called “sanitary landfills,” collect and treat their liquid effluents and also capture and use the generated biogas (about 50% methane). The next class of landfills burn their biogas in special flares to avoid the emission of methane to the atmosphere. At the very bottom of the waste management hierarchy are the age-old landfills that make no effort to collect their liquid and gas emissions and, regrettably, dot the surface of the Earth.

It should be kept in mind that, as in the case of every rule, there can be notable exceptions to the hierarchy of waste management. For example, as mentioned earlier, some yard wastes are used in California as Alternative Daily Cover (ADC) in sanitary landfills; in fact an LCA study by Columbia University (Rob van Haaren thesis) showed that the ADC use of yard wastes had a lower carbon footprint than composting, because of the avoidance of digging up and using soil as daily cover.

Greenhouse Gas Effects of Waste Management

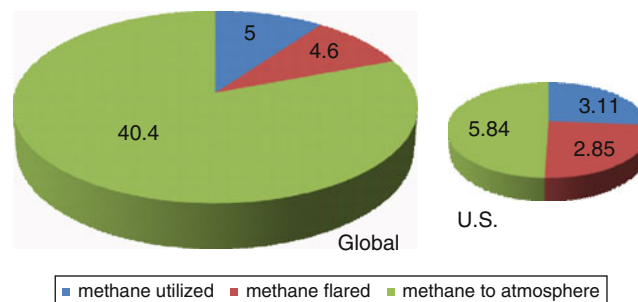
Landfilling is presently a major anthropogenic source of methane, a gas that is considered to be the second

most important of the greenhouse gases that are affecting climate change. The only two options for decreasing the emissions of methane in landfill gas (LFG), estimated presently to be 850 million tons of carbon dioxide, are accelerated growth of WTE capacity and increased LFG capture in the interim period. The global waste-to-energy capacity is about 190 million tons and at the current rate of growth of the WTE industry, about 2% annually, it cannot catch up with the projected growth of MSW generation. Therefore, it is necessary for governments to encourage WTE growth and also ensure that new and remaining landfills are constructed and operated so as to avoid contamination of ground and surface waters and maximize the capture and utilization of methane in landfill gas.

USA is the world’s largest landfiller with about 225 million metric tons, followed by China with an estimated 200 million metric tons. However, the US landfilling industry and the USEPA Landfill Methane Outreach Project have made a determined effort to collect much of the landfill gas (LFG). [Table 4](#) and [Fig. 5](#) show the generation and capture of LFG

Waste-to-Energy, Introduction. Table 4 US and global generation and capture of LFG (in million metric tons)

	LFG methane generation	LFG methane utilization	LFG methane flaring	LFG fugitive methane	CO ₂ equivalent
USA landfills	11.8	3.11	2.85	5.84	123
Global landfills	50	5	4.6	40.4	848
USA as % of global	23.6	62.2	62.0	14.5	14.5



Waste-to-Energy, Introduction. Figure 5

Global and US. generation and capture of landfill methane (all numbers are in million tons per year)

(50% methane) in the USA and worldwide, as estimated by the author in a 2008 publication (GWMS).

Thermal Treatment Technologies

As stated earlier, the only proven alternative to landfilling mixed or post-recycling MSW is combustion or gasification to recover electricity, heat, syngas, and metals. Worldwide, there are over 800 thermal treatment plants, most of them in EU, Japan, and the USA.

Grate Combustion of "As-Received" MSW

In a mass burn WTE, the MSW bags and wastes are discharged from the collection vehicles into the waste bunker in a fully enclosed building. An overhead claw crane loads the solids in the charging hopper and a ram feeder at the bottom of the hopper pushes the solids on the moving grate (Fig. 6), which usually is inclined; the force of gravity and the periodic motion of the grate propel slowly the bed of wastes through the

combustion chamber. The grate is 3–6 m wide, depending on the furnace capacity, and about 8–10 m long; the average residence time of solids on the grate is about 1 h.

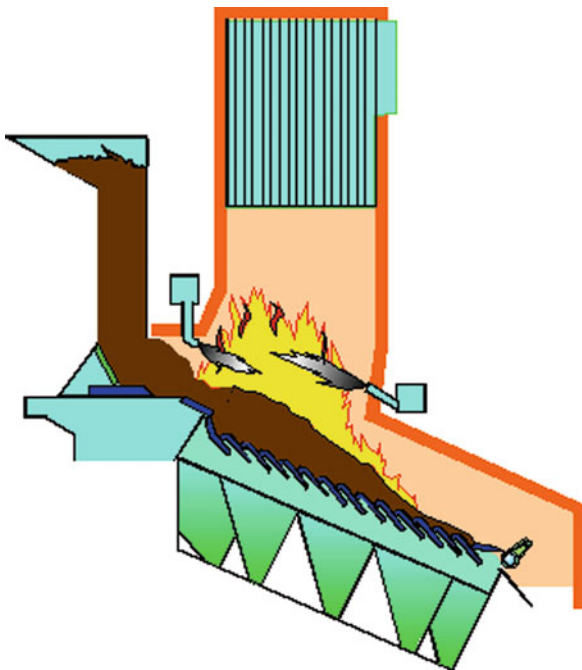
As will be described in detail in other entries in the WTE section, the heat contained in the combustion gases is transferred, through the furnace water wall and superheater tubes, to the high-pressure steam that drives the turbine generator. The low-pressure steam from the generator exhaust is condensed or, preferably, used for district or industrial heating. The most efficient waste-to-energy (WTE) facilities are co-generators of electricity (>600 kWh) and district heating (>500 kWh) per ton of MSW processed. As an example, over 30% of the district heating needs of Denmark are provided by 28 WTE plants that serve this nation of five million people.

The mass burn WTE technology has the advantage of the cumulated experience of several decades and this has resulted in very high plant availability, over 8,000 h of operation per year. It is also the simplest way to thermally treat MSW because it requires little contact with the MSW and very few people to operate and maintain. A plant processing one million tons of MSW can be operated by fewer than 60 people. However, they are very costly to build.

Most WTE plants built in the last decade are of the grate combustion. A 2006 EEC survey of three dominant technologies (Martin, Von Roll, Keppel-Seghers) showed consistent growth of about three million tons of mass burn capacity per year. Other technologies, such as direct smelting (JFE, Nippon Steel), fluid bed (Ebara), and circulating fluid bed (Zhejiang University, China) have accounted for an additional estimated growth of another million tons per year.

Combustion of Refuse-Derived Fuel (RDF)

The MSW feedstock of WTE processes has a typical heating value of 10 MJ/kg. Whatever is done to decrease the MSW particle size (e.g., shredding), increase its calorific value (e.g., drying and sorting of inorganic constituents), or add auxiliary energy in the combustion chamber (e.g., oxygen enrichment or use of thermal plasmas) requires the use of electrical energy that must be justified by increased energy production per ton of MSW, or higher productivity per unit



Waste-to-Energy, Introduction. Figure 6
WTE combustion chamber with inclined moving grate

volume of furnace and boiler that results in lower capital investment costs.

The Refuse-Derived Fuel (RDF) technology represents the simplest possible advance over grate combustion. The original RDF concept was to allow separation of marketable materials from the MSW feed before it enters the combustion chamber. Basically, this process consists of single-shredding of the MSW, sorting out some of recyclable materials, and then combusting the resulting RDF. This process is used mostly in the USA where there are 12 WTE facilities, ranging in capacity from 360 to 2,700 t per day and in total process about six million metric tons of MSW annually. A variation of the RDF process is the Mechanical Biological Treatment (MBT) process used in Europe that converts MSW to a solid recovered fuel (SRF) that can be co-combusted with coal in cement and also power plants, provided that their air pollution control systems can handle the resulting process gas.

Obviously, from a reaction engineering perspective, shredding of the highly heterogeneous MSW to a more uniform particle size and composition should be beneficial: Heat and mass transfer rates increase with smaller particle size; also, a certain degree of homogenization should facilitate combustion control in the furnace. Since the drying, volatilization, and combustion rates would be higher, the specific productivity (e.g., tons per unit volume of combustion chamber) of shredded WTE should also be higher than for a mass burn WTE. However, for wider adoption of the RDF technology, these advantages must translate into lower capital and operating costs than the grate combustion technology.

Emerging WTE Technologies

It should be stressed that while the novel WTE processes are called “gasification,” they are in fact a combination of partial oxidation and gasification, followed either by combustion of the volatiles and steam generation, similarly to grate combustion, or using the syngas in an engine or turbine generator, or converting to a synthetic fuel.

As in the case of the RDF and MBT processes, gasification technologies need to demonstrate what are, or will be, the specific advantages of more complex systems over the constantly evolving grate combustion

technology. It cannot be better control of emissions because, since the 1990s, all WTE plants have reduced their emissions to practically insignificant levels. It may be higher energy recovery because a gas turbine has higher thermal efficiency than the steam turbine used in grate combustion plants. There is a definite advantage of most gasification processes in vitrifying the ash residue. Also, a lower capital investment per ton of capacity than for grate combustion plants would be a definite advantage, especially for small city applications.

Beneficial Use of WTE Ash

The residues of waste-to-energy facilities are bottom ash, about 20–25% of the MSW by weight, and “fly” ash collected in the Air Pollution Control (APC) system, about 2–3% of the MSW. As the APC systems of WTE facilities have improved greatly, the captured heavy metals, dioxins, and other undesirable contaminants end up in the fly ash. At the present time, most of the US WTE facilities mix bottom and fly ash to form a “combined” ash that is chemically inert and is used for landfill maintenance and daily cover, in place of the 15-cm soil cover required by USEPA. Since the USA is the world’s largest landfiller, there is a great need for use of WTE ash in landfills. However, since there are no alternative beneficial uses of WTE ash outside landfills, the WTE companies do not get much benefit from supplying it to landfills. In fact, its disposal for landfill maintenance represents a substantial operating cost.

Bottom ash does not contain dioxins and volatile metals and its chlorine and sulfur concentrations are very low. It can be used beneficially in road construction, remediation of extinct mines, and other uses, as is done in several other nations. Numerous demonstration programs, in the USA and abroad, have proven scientifically that bottom ash can be processed to generate an engineered and environmentally safe aggregate for diverse construction applications. In fact, the WTE facility of AEB Amsterdam (1.5 million tons of MSW annually) processes its ash in a novel way so that only 1% by weight of the MSW combusted has to be landfilled.

Need for Government Support

Government support for new WTEs in the USA, China, and other landfilling nations will increase the

rate of growth of WTE but, for several decades, humanity will continue to depend on landfilling. Eventually, only inorganic, non-recyclable materials will be landfilled, as is already done in several nations. However, until there is sufficient WTE capacity, there is much to be done with existing and new landfills. In particular, it is necessary for developing nations like China and India to follow the example of USA in constructing sanitary landfills that prevent liquid effluents from contaminating ground and surface waters and reduce methane and other gaseous emissions to the atmosphere.

Management of solid wastes has become as important in our times as the supply of water and the treatment of wastewater. All three are part of the Commons that is usually neglected by individuals but must be taken care by the community and the nation.

Bibliography

1. Themelis NJ, Kim YH, Brady MH (2002) Energy recovery from New York City municipal solid wastes, *Waste Management & Research* 20(3):223–233
2. EPA report (2011) GHG emissions of solid wastes. (http://www.epa.gov/climatechange/emissions/downloads/08_Waste.pdf)
3. Themelis NJ, Ulloa PA (2007) Methane generation in landfills. *J Renew Energy* 32(7):1243–1257
4. Themelis NJ, Ulloa PA (2005) Capture and utilization of landfill gas. In: *Renewable energy 2005*, pp 77–81. www.sovereign-publications.com/renewable-energy2005-art.htm
5. Matthews E, Themelis NJ (2007) Potential for reducing global methane emissions from landfills. In: *Proceedings Sardinia 2007, 11th international waste management and landfill symposium*, Cagliari, 1–5 Oct 2007, pp 2000–2030
6. USEPA, Inventory of U.S. greenhouse gas emissions and sinks: 1990–2006. (2006) www.epa.gov/climatechange/emissions/downloads/08_Waste.pdf
7. USEPA-LMOP (2010) Energy projects and candidate landfills. www.epa.gov/lmop/proj/index.htm
8. Themelis NJ (2007) Global growth of traditional and novel thermal treatment technologies. *Waste Management World, Review Issue 2007–2008*, pp 37–47, July–Aug 2007
9. HSC Chemistry (2011) Chemical reaction and equilibrium software. Outotec, Finland. <http://www.hsc-chemistry.net/>
10. Lee S-H, Castaldi M (2009) Effects of hydrogen chloride gas concentrations on corrosion rates of tube alloys under WTE environments. In: *16th annual North American waste-to-energy conference*, NAWTEC 16. www.wtert.org/sofos
11. Kaufman SM, Themelis NJ (2009) Direct method to characterize and measure flows of municipal solid waste in the United States. *J Air Waste Manage Assoc* 59: 1386–1390

Waste-to-Energy: Energy Resource in Solid Wastes

DIETER O. REIMANN

Confederation of European Waste-to-Energy Plants (CEWEP), Bamberg, Germany

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Recovery of Usable Energy from Waste

Potential for Increasing Energy Recovery of WTE Plants

Future Directions

Bibliography

Glossary

APC residues Residues collected in Air Pollution Control system of a plant.

Best available technology (BAT) The most effective methods for achieving a high level of protection for the environment; developed on a scale that allows implementation under technically and economically viable conditions (in the USA, called Maximum Achievable Control Technology: MACT).

BREF Reference Document on the Best Available Techniques (e.g., BREF Waste Incineration (WI) and BREF Energy Efficiency (ENE)).

Energy equivalence factors As per definition of the R1 formula, electricity produced must be multiplied by a factor of 2.6; heat produced must be multiplied by a factor of 1.1.

MSW incineration (MSWI also called waste-to-energy or WTE) a facility that includes waste storage bunker, waste feeding equipment to the furnace, boiler, flue gas cleaning system, and energy recovery system (steam turbine, etc.).

Municipal solid waste (MSW) Waste from households as well as commercial, industrial, and institutional waste that, by its nature and composition, is similar to waste from households. This includes, e.g., bulky waste, residual waste after recycling operations, RDF-refuse derived fuels mainly from

municipal waste. The categories of waste that may be treated in a WTE plant are listed in the permit by the competent authority.

Produced energy Total energy generated as electricity and heat for sale and use within the plant.

R1 energy efficiency factor Determines whether or not a WTE plant can be classified as a recovery operation, under EU legislation.

R1 formula Used to calculate the R1 factor under EU legislation.

Sankey diagram Flow diagram in which the width of the arrows is proportional to the flow quantity.

Waste incineration Thermal treatment of MSW in an environmentally sound way and recover its energy content as electricity, heat, or combined heat and power.

Definition of the Subject and Its Importance

In waste legislation and policy, the waste management hierarchy is a list of waste management methods, according to what constitutes the best overall environmental option. It consists of: prevention, reuse, recycling, composting, energy recovery, and disposal [1]. Waste incineration plants are classified either as “recovery,” or as “disposal” operations. This depends on the fraction of energy recovered from that contained in the waste. It is important to note that “recovery” is defined as any operation through which wastes replace other materials that would be needed to fulfill a particular function in the plant or in the wider economy [1].

An appropriate method to assess whether or not a waste incineration plant is a recovery operation is to evaluate the energy produced and used from the waste. The European Commission has quantified this distinction by means of the R1 energy efficiency formula. A plant can only be classified as a recovery operation when its R1 factor is at least 0.60 for existing, and 0.65 for new plants. In the R1 formula this is determined using equivalence factors; a similar method is also used in Switzerland, but without using equivalence factors.

The R1 factor provides a major incentive for WTE plants to improve their energy efficiency. High R1 factors increase the contribution WTE plants make to the national energy supply, saving primary fossil fuel resources, e.g., in power or heating plants, and reducing the emission of greenhouse gases. The R1 factor

also serves as a driving agent in the competition among operators of WTE plants to reach a higher ranking than other facilities.

Introduction

The three main purposes of MSWI or WTE are to reduce the volume of waste by means of thermal treatment; control effectively the emissions of this operation; and recover as much as possible of the energy content of the wastes.

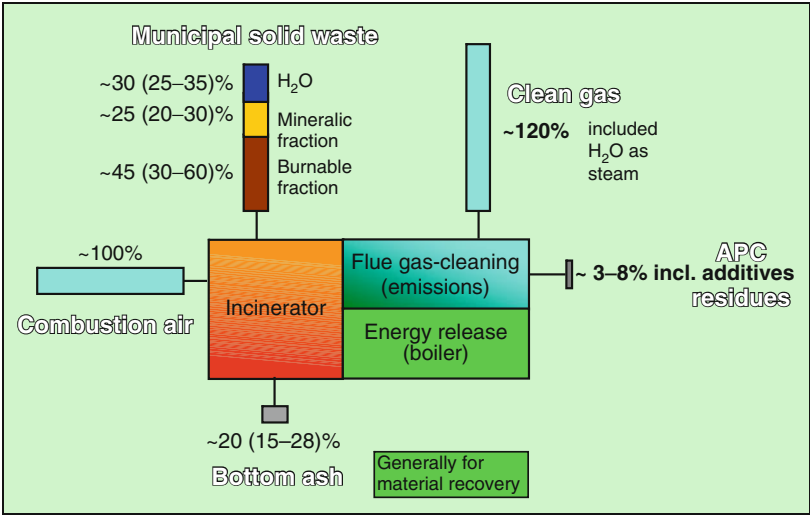
Material Flow

The aim of the incineration process is to reduce the quantity of waste and produce as many recoverable materials from the residues as possible, so that only a very small part need to be deposited in a landfill. The technologies used to achieve the material flow are shown in Fig. 1 [5, 10]. They primarily depend on the type of grate or other combustion device, the composition and calorific value of the waste to be incinerated, the operating conditions in the furnace (e.g., % excess air), and the combustion temperature.

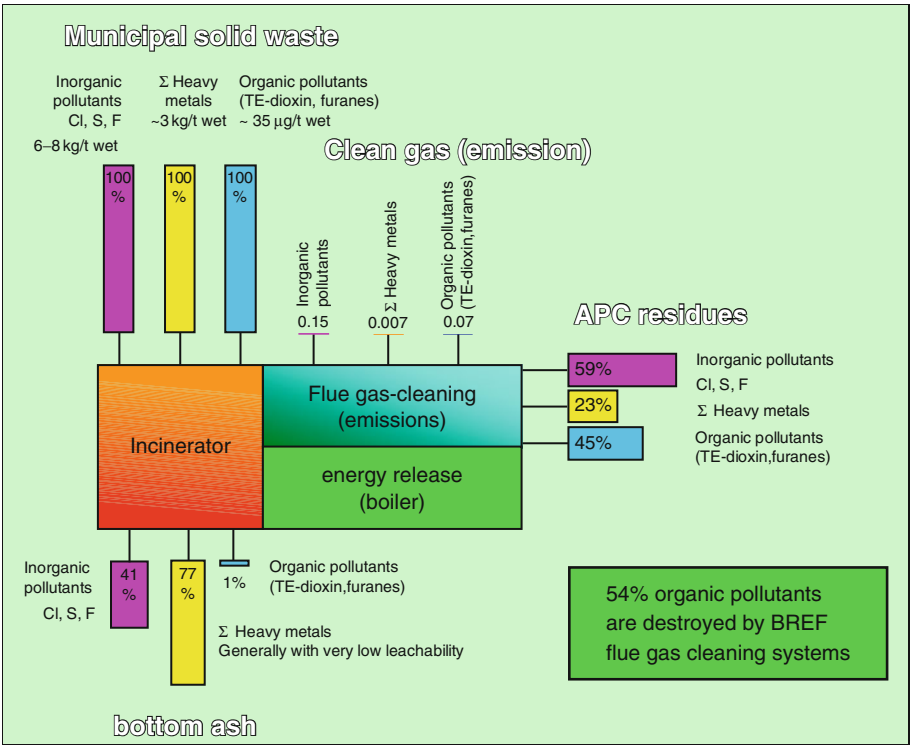
Pollutants and Emissions

One of the aims of the incineration process is to destroy organic pollutants and to separate heavy metals and inorganic pollutants that are contained in the incoming waste. By employing highly efficient flue gas cleaning technology, WTE plants can ensure that emissions into the ambient air are negligible. The data in Fig. 2 for the input and fate of pollutants in a WTE plant are from a life cycle assessment (LCA) study [2]. The destruction of organic pollutants takes place in the combustion chamber and in the first pass of the boiler, maintained at a temperature above 850°C; this is followed by adsorption of the remaining organic pollutants followed in the APC system, by means of the injection of activated carbon, or other method.

The heavy metals and other inorganic pollutants in the MSW are depending on their volatility volatilized during combustion. The higher the bed temperature, which should not exceed the melting point of bottom ash, the fewer pollutants will remain in the bottom ash. This is an important quality criterion for beneficial use of the bottom ash. The volatilized metals and inorganic



Waste-to-Energy: Energy Resource in Solid Wastes. Figure 1
MSWI: Material flow diagram showing range of MSW composition and distribution of MSW weight as flue gas, bottom ash, and APC residues



Waste-to-Energy: Energy Resource in Solid Wastes. Figure 2
MSWI: Path and percent distribution of pollutants contained in MSW processed in a WTE plant in Europe [2]

pollutants are then captured in the Air Pollution Control system, as described in detail in other chapters of this Encyclopedia. As a result, WTE plants designed and operated according to the Waste Incineration BREF [3], can meet the very stringent EU emission standards for waste incineration plants [4].

Energy from Waste

A waste incineration plant should aim to recover as much as possible of the energy available in the MSW feed. Figure 3 shows how the chemical energy stored in the waste is distributed in a WTE plant. The MSW feed into a WTE plant is a mixture of several fractions with very different calorific values, ranging from 1.6 to 4.3 MWh/t and an average value for Europe of 2.8 MWh (about 10 GJ) per ton of “as received” MSW (Fig. 3). An appropriate calculation method, based on the main data of a plant, is described in Section Net Calorific Value of MSW (NCV; Also Called Low Heating Value or LHV).

In the waste bunker, the different components of the MSW are mixed by the overhead crane, to ensure optimal combustion within the furnace and avoid the need to use auxiliary fuel, except during start-up or shutdown of a unit. The importance of the combustion

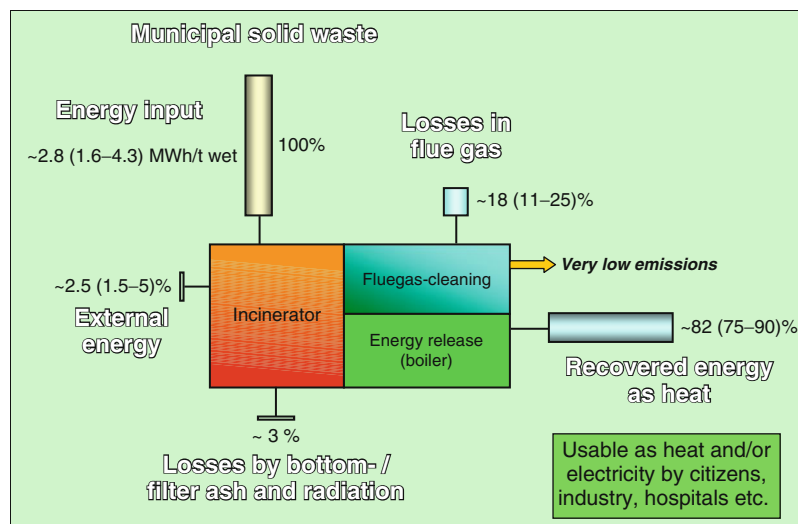
control system within the furnace is discussed in detail in Waste to Energy (WTE): The AE&E Technology and The Martin WTE Technology entries of this Encyclopedia.

Projected Energy Losses of the Incineration Facility

The heat loss in the bottom ash, the convection and radiation losses of furnace and boiler, and the loss associated with incomplete burnout of carbon in the ash correspond to about 3% of the energy input in the MSW feed. In the R1 formula, they are taken into account by the fixed correction factor of 0.97. This is the reason why only 97% of the heating value of MSW (plus any auxiliary heat used in the plants) is entered into the denominator of BREF WI [3].

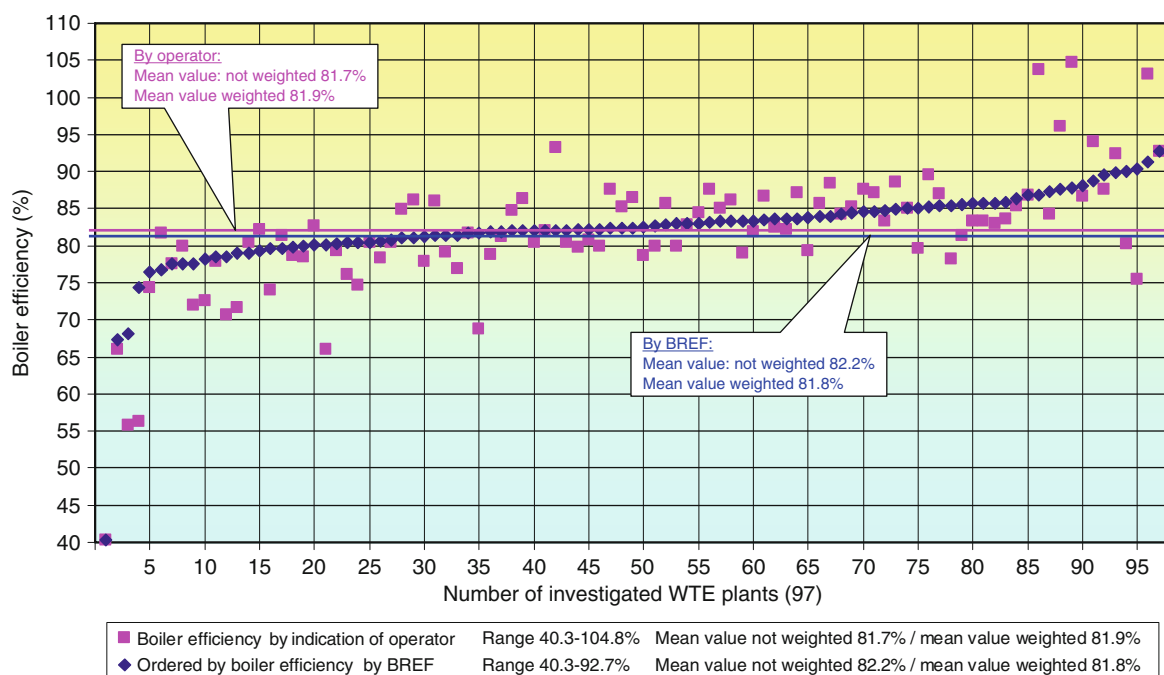
Boiler Efficiency and Flue Gas Temperature After Boiler

The temperature in the flue gas leaving the boiler affects the energy recovery by the boiler. It is independent of the heating value of the MSW and in state-of-the-art systems is in the range of 180–230°C. This range is associated with steam parameters of 40 bar/400°C and results in a mean boiler efficiency of 82% [8]; in other words, 18% of the input energy in the MSW is lost in the flue gas.



Waste-to-Energy: Energy Resource in Solid Wastes. Figure 3

MSWI: Flow diagram and distribution of energy contained in MSW processed in a WTE plant



Waste-to-Energy: Energy Resource in Solid Wastes. Figure 4

Boiler efficiencies of 97 European WTE facilities (CEWEP 2006)

Figure 4 shows the calculated boiler efficiencies of 97 European WTE plants, investigated by the Confederation of European WTE Plants [8], using the formula:

% boiler efficiency = Energy contained in steam generated/0.97 (Energy contained in MSW feed + energy in any auxiliary fuel used in steam generation). The red squares show the actual data provided by the plant operators (2001–2004 data) and the blue triangles the calculated data according to the BREF WI formula.

It can be seen that many of the boiler efficiencies reported by the operators differ considerably from the BREF formula. This can be due to several factors, such as low or high estimated of heating values of MSW, inaccurate measurements of steam rates, etc.

When the steam energy is to be transformed to electricity, high steam parameters are required. In this case, relatively higher temperatures (up to 450°C) and pressures (up to 100 bar) of the steam are technically

possible, but the surfaces of the boiler membrane walls, at least in the first and second pass and the superheater tubes must be protected against corrosion and abrasion, by cladding or other means.

Modern incineration plants reach boiler efficiencies over 85%. Even higher efficiencies are achievable with wastes containing low chlorine than MSW, which contains 0.4–0.6% per tonne of wet MSW.

Net Calorific Value of MSW (NCV; Also Called Low Heating Value or LHV)

Many NCV (or LHV) data reported by the operators show a large deviation from the correct values of NCV. Because of this, a NCV formula has been defined by the BREF WI that in effect uses the furnace and boiler as a giant calorimeter. It is based on a simple energy balance between the energy inputs and the output of a WTE plant:

Formula 1: Formula for calculation of the net calorific value (NCV; or LHV in the USA) of MSW feed, according to BREF Waste Incineration (WI) using measured or calculated data:

$NCV = (1.133 \times (m_{st\ w}/m) \times c_{st\ x} + 0.008 \cdot T_b) / 1.085$, in GJ/t wet MSW
where
NCV = net (or lower) calorific value (NCV) of MSW, in GJ/t (wet)
$m_{st\ w} = m_{st\ x} - (m_f \times (c_f / c_{st\ x}) \times \eta_b)$, in t steam/t wet MSW
and
$m_{st\ w}$ = mass flow of steam produced from waste combustion, in t/year, or t/day
m_f = mass flow of steam from auxiliary fuel combustion, in t/year, or t/day
$m_{st\ x}$ = total amount of the steam produced, in t/year, or t/day
m = mass of waste combusted, in t/year, or t/day
$c_{st\ x}$ = net enthalpy of steam (i.e., minus enthalpy of boiler water), in GJ/t wet
c_f = net heating value (LHV) of auxiliary fuel used for steam production, in GJ/t
T_b = temperature of flue gas after boiler (at 4–12% O ₂) – T_{input} of air, in °C
0.008 = specific sensible heat capacity of flue gas, GJ/t wet/°C
1.133 and 1.085 = constants derived from regression equations
η_b = efficiency of gas-to-steam heat transfer in the boiler (approx. 0.80)

Because the greatest part of the energy output goes to the steam exiting the boiler ($m_{st\ x} \cdot c_{st\ x}$), accurate measurement of the steam rate is very important in using the above formula to calculate the NCV (or LHV) of the waste. An example of “double counting” errors in the steam rate is steam used to preheat the combustion air. On the other hand, steam quantities that should be included in the steam rate of the plant are steam extracted ahead of the steam measuring device for heating the flue gases (e.g., before SCR, after wet scrubbers, before fabric filters) or injection of water for continuous boiler cleaning, or in the form of NH₄OH for NO_x reduction; the energy needed to evaporate these sources of water is not counted by the steam measuring device.

An example of NCV (or LHV) calculation, based on the NCV formula of BREF WI, is shown in Table 1 for a WTE plant with the annual energy flows shown in Table 1 [11, 12].

The calculations made in Table 1 show that by taking into account all relevant energy flows in this WTE plant, the calculated NCV (LHV) of the waste was 10.0 GJ/t of MSW at a boiler efficiency of 85.3%. In comparison, for a simplified NCV calculation, using only the energy of measured steam minus imported energy, the NCV would increase to 10.3 GJ/t and the boiler efficiency would decrease to 83.4%.

To get accurate and comparable NCVs of the waste combusted at WTE plants of different configurations, it is necessary to collect the complete energy flows from each plant. Figure 5 shows the results of an investigation that compared the net calorific values (NCV or LHV) of 97 European WTE plants, as calculated by the BREF WI formula (blue triangles) and reported by the plant operators (red squares; data for 2001–2004). A similar variation between these two values is observed as in Fig. 4 that compared calculated and reported boiler efficiencies.

The observed differences in calorific value among different European WTE plants may be due to several factors, including the prevailing rate of recycling, the degree of industrialization, and the economic development of the community served by a WTE. Figure 6 classifies NCV data from 231 European WTE plants into three regions: Northern, middle, and southern Europe [9].

The reason for the relatively wide range of calorific values of MSW combusted at different WTE within a region is most likely due to different fractions of high heating value wastes treated at different plants (e.g., industrial and commercial waste, nonrecyclable wrapping, and waste wood). For example, in northern Europe which is the coldest climate zone, high calorific waste is used to complement the available MSW and provide the demand for district heating and reduce the consumption of fossil fuels.

The low heating values in southern Europe are most likely due to the incineration of relatively larger quantities of food and green wastes and smaller quantities of plastics and paper.

Recovery of Usable Energy from Waste

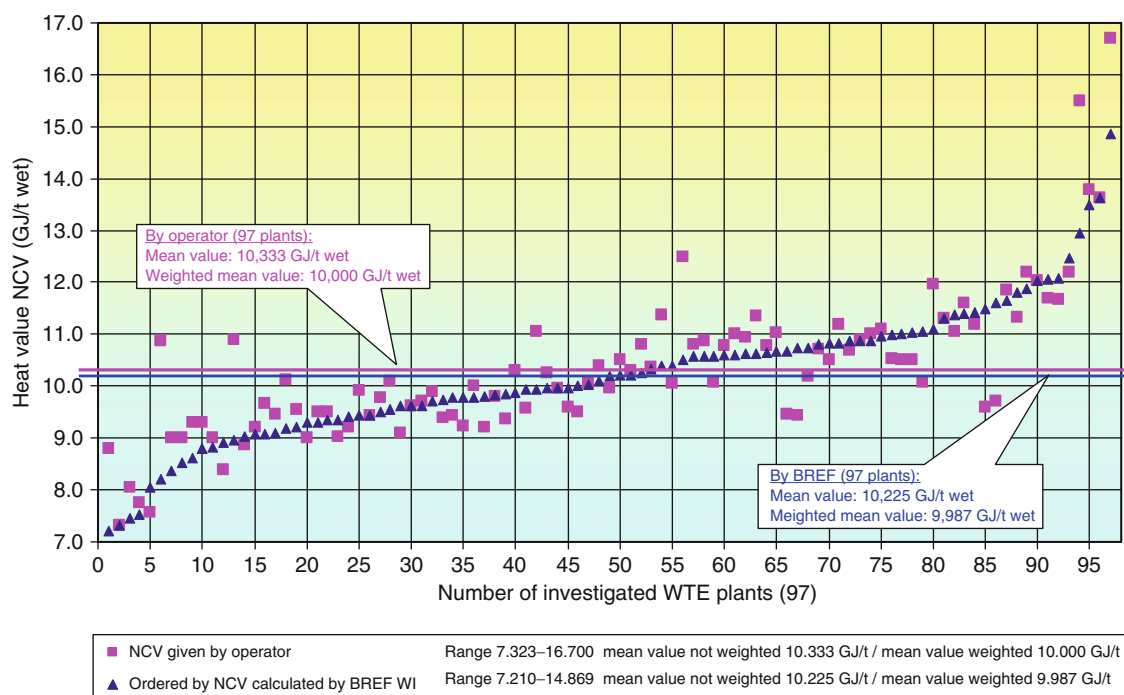
The potential for energy recovery from WTE depends not only on the combustion technology used, but also

Waste-to-Energy: Energy Resource in Solid Wastes. Table 1 Example of calculation of NCV (or LHV) using the BREF WI formula

A	Boiler efficiency $(1,738,100 + 15,992)/(0.97 * (210,000 * 10.038 + 12,100)) * 100 = 85.3\%$	η_b	85.3	%
B.1	Municipal waste incinerated	m_w	210,000	t/year
B.2	Measured steam quantity after boiler with the main measuring device	$m_{st \ x}$	630,000	t/year
B.3	Net enthalpy of steam 40 bar/400°C/130°C boiler water	$C_{st} (3,214 + 544)$	2,670.000	kJ/kg
B.4	Delivered steam to a third party with a second measuring device	m_w	20,000	t/year
B.5	Net enthalpy of steam 45 bar/257°C with no backflow of condensate	$C_{st} \times (2,800)$	2,800	kJ/kg
B	Total energy output based on measured steam quantities	$m_{st \ x} \times C_{st \ x}$	1,738,100	GJ/year
C.1	Light fuel oil for start-ups and shutdowns	m_f steam producing	120	t/year
C.2	Light fuel oil for keeping combustion temperature $>850^\circ\text{C}$	m_f steam producing	215	t/year
C.3	Sum of used light fuel oil	m_f steam producing	335	t/year
C.4	Energy input by fuel	$m_f \times c_f (42,73 \text{ GJ/t})$	12,100	GJ/year
C	Included in total energy output from measured stream	$\eta_b \times m_f \times c_f$	10,321	GJ/year
D.1-1	Steam for soot blowing and for NH_4OH injection SNCR (both extracted after total steam measuring device)	$m_{st \ x}$	12,500	t/year
D.1-2	Net enthalpy of steam 40 bar/400°C/0.2 bar/220°C steam in flue gas	$C_{st} \times (3,214 - 2,918)$	296	KJ/kg
D.1	Energy by double measured steam quantity for soot blowing and NH_4OH injection of SNCR	$m_{st \ x} \times C_{st \ x}$	3,700	GJ/year
D.2-1	Steam for heating up primary combustion air from 20°C to 130°C (extracted after steam measuring device)	$m_{st \ x}$	37,600	t/year
D.2-2	Net enthalpy of steam 3 bar/144°C/100°C in condensate	$C_{st \ x} (2,748 - 419)$	2,329	kJ/kg
D.2-3	Energy input by steam	$m_{st \ x} \times C_{st}$	87,570	GJ/year
D.2	Energy by double measured steam quantity for heating up primary combustion air from 20°C to 130°C	$\eta_b \times m_{st \ x} \times C_{st \ x}$	74,698	GJ/year
D.3-1	Liquid quantity as water and NH_4OH injected before the total steam measuring device	m_{liquid}	4,600	t/year
D.3-2	Net enthalpy as steam 0.2 bar/220°C in flue gas	$C_{st} \times (2,918)$	2,918	KJ/kg
D.3	Not measured energy by injected liquids into the combustion chamber before steam measuring device	$m_{liquid} \times C_{st \text{ flue gas}}$	13,423	GJ/year

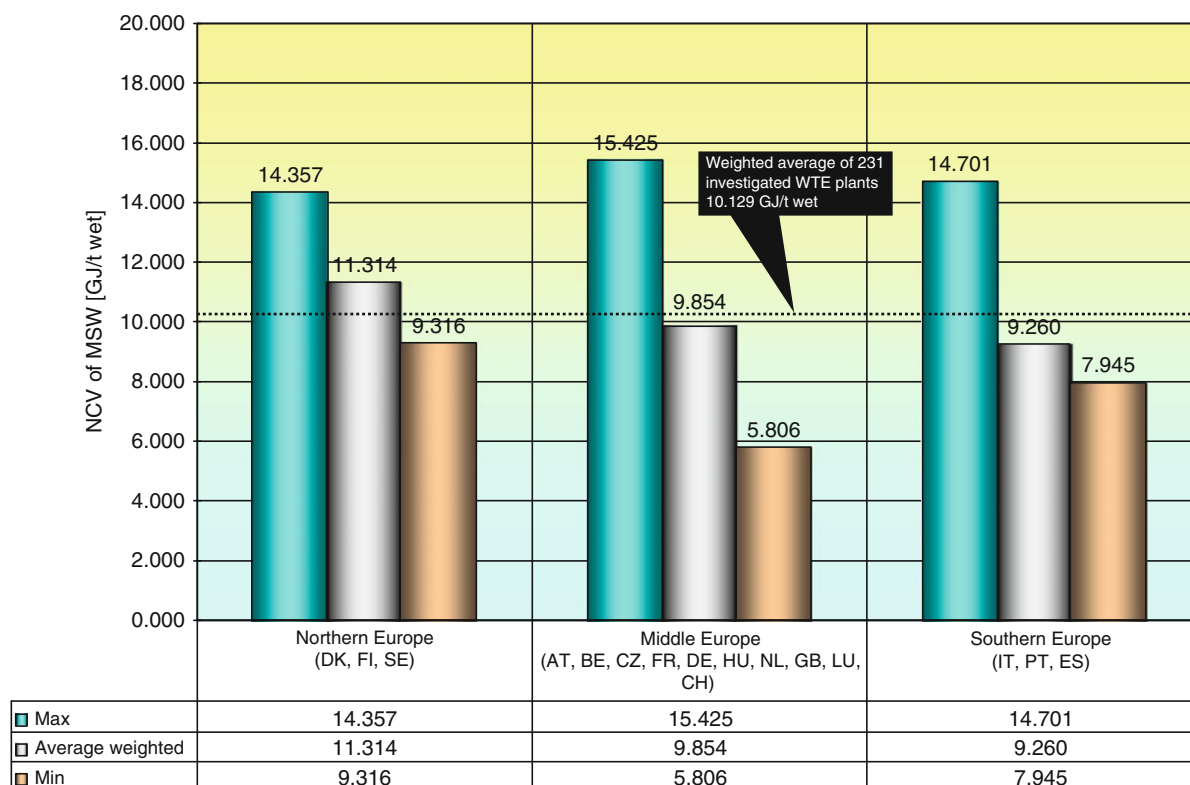
Waste-to-Energy: Energy Resource in Solid Wastes. Table 1 (Continued)

D.4-1	Steam extracted before the steam measuring device for heating up flue gas from 110°C to 120°C	$m_{st} \times$	6,600	t/year
D.4-2	Net enthalpy of steam 45 bar/257°C/90°C condensate	$C_{st} \times (2,800 - 377)$	2,423	KJ/kg
D.4	Not measured energy for heating up flue gas from 110°C to 120°C with steam out of the boiler drum	$m_{st} \times C_{st}$	15,992	GJ/year
D.5-1	Energy extracted by grate cooling before the steam measuring device for heating up the backflow of district heat		3,800	MWh/year
D.5-2	Transferred from MWh into GJ		3.6	GJ/MWh
D.5	Not measured energy from grate cooling for heating up the backflow of district heat		13,680	GJ/year
D	Sum of double or not measured energies included in total energy output from measured steam	$\eta_b \times m_f \times C_f$	-35,303	GJ/year
$\Sigma (B - C + D)$: NCV relevant energy flows of this example			1,692,476	GJ/year
NCV determination by BREF WI formula based on the investigated, NCV relevant energy flows ($\eta_b = 85.3\%$)			10.038	GJ/t MSW
NCV determination by BREF WI formula based only on the energy from the measured steam in B ($\eta_b = 83.4\%$)			10.265	GJ/MSW



Waste-to-Energy: Energy Resource in Solid Wastes. Figure 5

Net calorific values (or LHV) of 97 European WTE plants (CEWEP 2006, 2009)



Waste-to-Energy: Energy Resource in Solid Wastes. Figure 6

Analysis of calorific values of MSW combusted in 231 WTE plants in Europe

on location and size of the plant, whether only electricity or heat are generated, and the incentives provided by government for the recovery of renewable energy and reduction of greenhouse emissions.

Status of Energy Recovery from Waste in Europe

Table 2 presents an overview of energy forms and average recovery rates of 231 European WTE plants, in 2004–2007. Forty-one plants produce only heat, 75 only electricity, and 115 plants both heat and power (CHP). The “specific” values denote energy recovered per ton of MSW; the percent values indicate the energy output (MWh of heat and/or electricity recovered) per MWh energy input in the MSW feed.

For WTE plants producing only heat, the energy recovery is over 70%, with a maximum of 96.8% for the most efficient plant (only hot water producing). Plants

producing electricity exhibit a thermal efficiency of over 20%, with a maximum of 32.4% for the best plant. However, it should be kept in mind that electricity is a more valuable source of energy than heat; The EU BREF multiplies MWh of electricity by the factor of about 2.6; therefore, the 20% mentioned above corresponds to 52% actual thermal efficiency.

In the case of WTE plants that supply both heat and power (CHP), the thermal efficiency is over 35% for heat plus over 14% for electricity generation; the maximum value reported was 89.6% for a WTE plant recovering heat combined with 1.2% for electricity generation.

Classifying European WTE Plants as Recovery or Disposal Operations

As noted earlier, European WTE plants can be classified as recovery or disposal operations, depending

Waste-to-Energy: Energy Resource in Solid Wastes. Table 2 Generation of electricity and/or heat by 231 WTE plants in Europe (2001–2004)

Kind of energy recovery		All investigated WTE plants not classified	%	WTE plants only heat producing	%	WTE plants only electricity producing	%	WTE plants CHP producing	1.%
Number of WTE plants	[–]	231	100	41	17.7	75	32.5	115	49.8
Total throughput of MSW	t wet/year	45,518,189	100	4,568,219	10.0	12,722,342	28.0	28,227,628	62.0
Specific energy input by MSW (without import) (NCV)	MWh abs./t wet	2.814	–	2.953	–	2.618	–	2.867	–
Total specific energy input (included import)	MWh abs./t wet	2.876	–	3.099	–	2.654	–	2.924	–
Total energy input (incl. import)	MWh abs./t wet	130,910,312	100	14,156,911	10.8	33,765,096	25.8	82,537,584	63.0
Specific electricity produced (Ep)	MWh abs./t wet	0.413	14.4	0	0.0	0.551	20.8	0.417	14.3
Specific electricity exported (Ep)	MWh abs./t wet	0.318	11.1	0	0.0	0.457	17.2	.306	10.5
Specific heat produced and used (Ep) ^a	MWh abs./t wet	0.905	32.2	2.210	71.3	0.092	3.5	1.067	36.5
Specific heat produced and used (Ep) ^b	MWh abs./t wet	1.202	41.8	2.486	80.2	0.399	15.0	1.356	46.4
Specific heat exported (Ep)	MWh abs./t wet	0.790	27.5	2.111	68.1	0.000	0	0.932	31.9
Specific total energy demand ((part of Ep) + Ef + Ei (th + el)) ^a	MWh abs./t wet	0.272	9.5	0.245	7.9	0.222	8.4	0.303	10.4
Specific total energy demand ((part of Ep) + Ef + Ei (th + el)) ^b	MWh abs./t wet	0.569	19.8	0.521	16.8	0.529	19.9	0.592	20.2
Including in total specific demand as Ef	MWh abs./t wet	0.015	0.5	0.018	0.6	0.010	0.4	0.009	0.3
Including in total specific demand as Ei(el)	MWh abs./t wet	0.019	0.7	0.101	3.3	0.004	0.2	0.017	0.6

Waste-to-Energy: Energy Resource in Solid Wastes. Table 2 (Continued)

Kind of energy recovery		All investigated WTE plants not classified	%	WTE plants only heat producing	%	WTE plants only electricity producing	%	WTE plants CHP producing	1.%
Including in total specific demand as Ei(th)	MWh abs./t wet	0.028	1.0	0.027	0.9	0.022	0.8	0.031	1.1
Including in total specific electricity produced as Ep (self-used electricity)	MWh abs./t wet	0.095	3.3	0.000	0.0	0.094	3.5	0.111	3.8
Including in total specific heat produced as Ep (self-used heat) ^a	MWh abs./t wet	0.115	4.0	0.099	3.2	0.092	3.5	0.135	4.6
Including in total specific heat produced as Ep (self-used heat) ^b	MWh abs./t wet	0.412	14.3	0.375	12.1	0.399	15.0	0.424	14.5

^aIn the real heat self used (demand) are, related to Sankey energy flow, beside the imported energy with 100% i.a. 18% energy self used for heating up of combustion air and all other processes with heat self demand, i.e., the steam to soot blowers, evaporation of APC residues, flue gas reheating, pipe heating, and building heating taken into account.

^bIn the heat self used (demand) related to the draft of the EU-Guidance for R1 formula, are additional to the load of "a" the total heat demand for boiler water heating up from an average temperature basis of 70°C to the individual boiler water temperature and 100% (instead of 18%) energy self used for heating up of combustion air included.

on their thermal efficiency [1]. The energy produced is calculated as the sum of energy exported plus energy used to operate the facility in a proper way, i.e., meeting at least all requirements specified by the permit of the plant.

Boundary Limits for Energy Balance

In order to calculate the energy efficiency of a plant, it is important to define the boundary limits of the plant, as illustrated in Fig. 7.

R1 Energy Efficiency Factor for the Member States of the European Union

The R1 energy efficiency factor is a nondimensional figure based on the first law of thermodynamics. Its purpose is to increase energy efficiency of WTE plants and minimize the demand of primary energy in the European Union [1]. This factor is calculated according to the following formula [6].

$$R1 = \text{Energy efficiency factor} = (Ep - (Ef + Ei)) / (0.97 \times (Ew + Ef))$$

where

Ep = energy produced per year as electricity or heat; electrical energy is multiplied by 2.6; heat produced for commercial use is multiplied by 1.1

Ef = energy input per year to the plant from fossil fuels

Ei = other external used per year in the plant

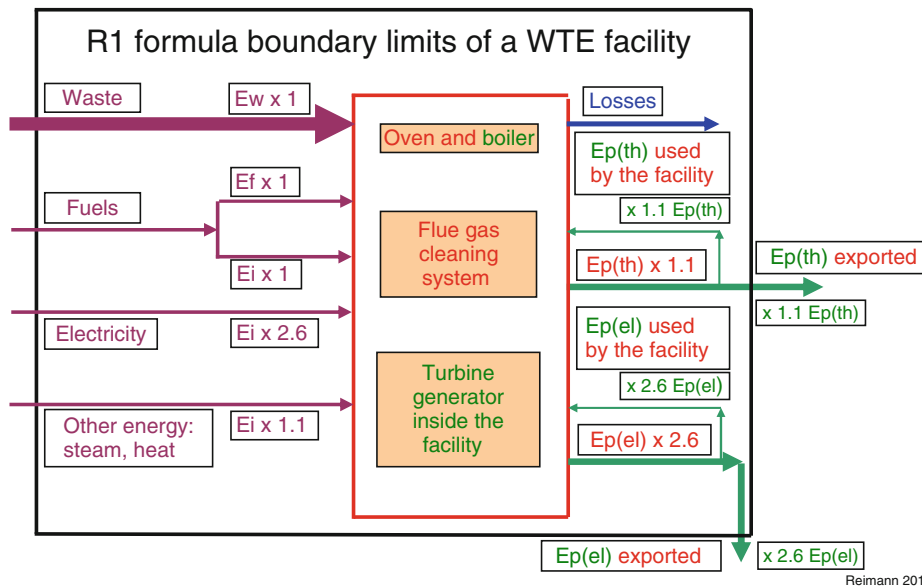
Ew = energy input per year to the plant from waste, on basis of its net calorific value (or LHV)

The factor 0.97 accounts for heat losses due to bottom ash and external radiation/convection

In the absence of external fossil fuels and electricity, formatting the R1 formula reduces to:

$$R1 = \text{Energy efficiency factor} = Ep / (0.97 \times Ew)$$

The equivalence factors for energy are based on savings of primary fuels in dedicated power and/or



Reimann 2010

Waste-to-Energy: Energy Resource in Solid Wastes. Figure 7

Boundary limits for the determination of the energy efficiency of a plant

heat plants. In the R1 formula, electrical energy is multiplied by the equivalence factor 2.6 and heat by the factor 1.1. To reach the “recovery” status (R1), the computed R1 factor, according to the above formula, must be equal to or higher than

- 0.60 for plants in operation and permitted, in accordance with applicable EU legislation, before January 1, 2009
- 0.65 for plants permitted after December 31, 2008

Figure 8 is based on the data presented in Table 2 for 231 WTE plants in Europe. It divides these plants into three classes: Plants producing only heat, producing only electricity, and producing both heat and electricity (2003–2007 data). For each class, the maximum, average and minimum R1 factors are shown. The averages values of energy efficiency are calculated in two ways: On basis of the R1 formula and from the Sankey energy balance, as discussed earlier.

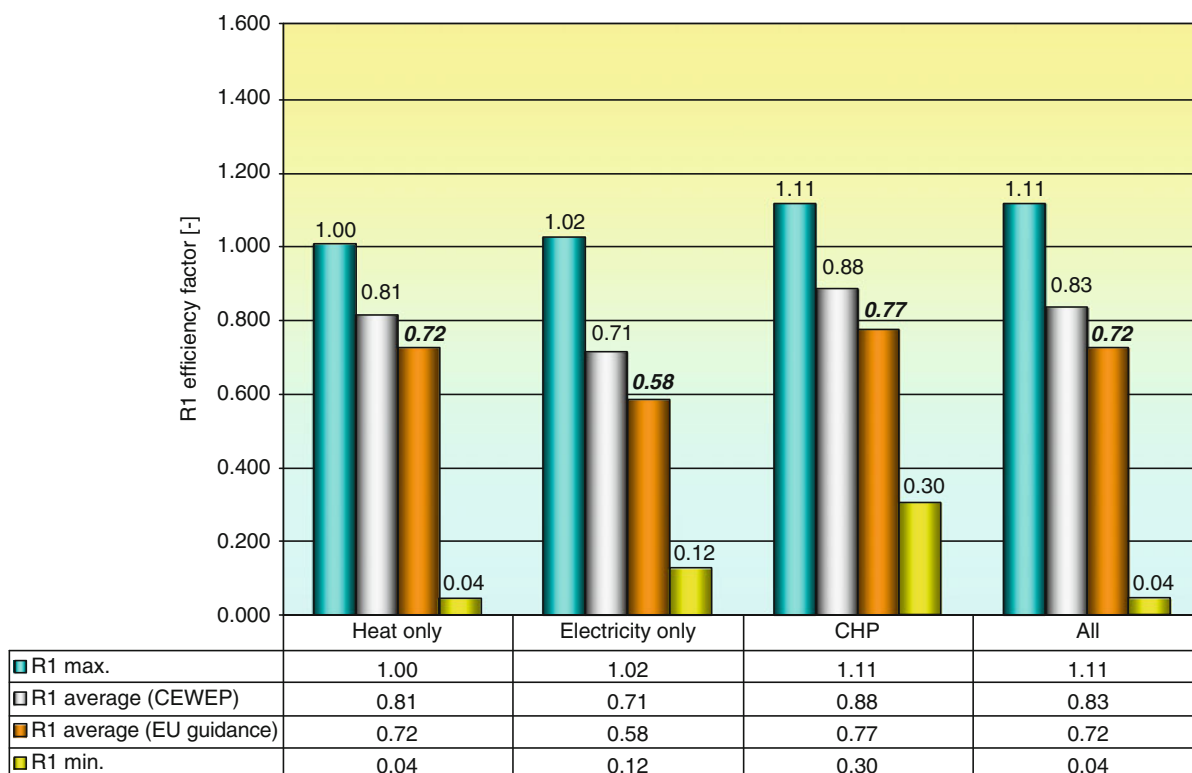
The highest R1 average based on CEWEP (0.88) was recorded for a CHP plant, producing both electricity and heat. The second highest average R1 factor was attained by a heat-generating WTE. The average R1 factor of WTE facilities generating only electricity was 0.71.

Potential for Increasing Energy Recovery of WTE Plants

Reimann [12] has suggested several ways to increase the energy efficiency of existing or new WTE facilities. For example, operators can aim to reduce excess air by optimal use of secondary air and recirculation of part of the flue gas stream. Other low cost measures are avoidance of fouling in boiler, ensuring a low flue gas temperature after boiler, reducing use of fossil fuel during start-up and shutdown, preheating the combustion air, and heating the flue gas for Selective Catalytic Reduction (SCR) by self produced steam instead of fossil fuels.

More costly measures recommended are frequent cleaning of heat exchanger tubes, providing for cooling of condenser by water sprinkling during high temperature excursions, increasing the boiler water temperature, use of a multistage turbine with reheating and dewatering of extracted steam in between the stages, and using boilers with higher steam parameters.

By far, the most effective means to increase the WTE thermal efficiency is by using the low-pressure steam for district heating or for providing heat to



¹⁾R1 calculation in accordance to Directive 2008/98/EC (WFD) 20/10/2008, ANNEX II, with equivalence factors: for electricity self produced and imported 1 MWh el=2.6 MWh el equ; for heat commercial used 1 MWh th=1.1 MWh th equ and according to BREF WI for imported fuel 1 MWh fuel=1.0 MWh fuel equ. Self used heat in Ep as described under ²⁾ in Table 2.

Waste-to-Energy: Energy Resource in Solid Wastes. Figure 8

Analysis of R1 energy efficiency factors of 231 WTE plants in Europe [9]

nearby industrial plants as base load, e.g., a paper recycling plant that requires a lot of heat. Of course, such measures are more applicable to new WTE plants and depend on the presence of customers for heat and the length of the heating or cooling period at a certain location.

Future Directions

Every WTE plant should strive not only to reach the R1 standard required for classification as a recovery plant but to exceed it. Higher recovery rates have several environmental and economic benefits, including the conservation of nonrenewable fuels and the reduction of greenhouse gas emissions with their climate change implications.

Bibliography

Primary Literature

1. European Parliament and Council (2008) Waste framework directive 2008/98/EC. Annex I and II. Off J L312:3–30, 22.11.2008. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008L0098:EN:NOT>
2. PE International and CEWEP (2008) Life cycle assessment of WTE plants in Europe: Study, Table 2.2. <http://www.cewep.eu/studies/climate-protection/art230,223.html> (available on request from CEWEP)
3. European Commission (2006) BREF waste incineration (WI) for Integrated pollution prevention and control (IPPC) EIPPC Bureau Sevilla. <http://eippcb.jrc.es/pages/FActivities.htm>
4. European Parliament and Council (2000) EU waste incineration directive 2000/76/EC. Off J L332:0091–0111, 28.12.2000. http://eur-lex.europa.eu/LexUriServ/site/en/oj/2000/l_332/l_33220001228en00910111.pdf

5. Reimann DO, Hämmerli H (1995) Verbrennungstechnik für Abfälle in Theorie und Praxis (Waste incineration in theory and practice). Schriftenreihe Umweltschutz, Bamberg p 80ff (out of print)
6. Reimann DO (2003) Ermittlung und Bedeutung von Kennzahlen zur Energie und Anlagennutzung sowie zu Wirkungsgraden für die Abfallverbrennung (Determination and meaning of identification numbers for energy and plants as well as efficiencies of waste incineration) Müll und Abfall, Heft 10:512–520
7. European Commission (2010) Draft of the EU-Guidance for R1 formula, Version April 2011
8. CEWEP (2006) CEWEP energy report (Status 2001–2004) by Reimann DO, p 5ff. <http://www.cewep.eu/studies/climate-protection/art230,223.html>
9. CEWEP (2009) CEWEP energy report II (Status 2004–2007) by Reimann DO, p 4ff. <http://www.cewep.eu/studies/climate-protection/art230,223.html>
10. Reimann DO (1991) Rostfeuerungen zur Abfallverbrennung (Grate fired combustion of waste). EF Verlag, Berlin. ISBN 3-924511-55-1
11. Reimann DO (2004) Energetische Bilanzierung von Müllverbrennungsanlagen (Balancing of energy from waste incineration plants). Erich Schmitt, Berlin, ISBN 3 503 08324 3
12. Reimann DO (2010) R1 as efficiency indicator: status quo and optimization potential. In: Second conference on biomass and waste combustion, Oslo. www.sintef.no

Books and Reviews

- Euroheat & Power (2005) District heating and cooling – Country by country/2005 Survey. Euroheat & Power, Belgium
- IPCC-Directive (2000) Good practice guidance and uncertainty management in National Greenhouse Gas Inventories – IPCC/OECD/IEA Programme. WMO Intergovernmental Panel on Climate Change UNEP, Bern, p 55ff

Waste-to-Energy: Fluidized Bed Technology

FRANZ P. NEUBACHER
 UV&P Environmental Engineering
 M.S. Chemical Engineering (T.U. Graz)
 M.S. Technology & Policy (M.I.T), Vienna, Austria

Article Outline

Glossary
 Definition of the Subject and Its Importance

Introduction

Technical Features and Limitations for Waste Treatment in FBC

Selected Concepts and Reference Projects for FBC Emissions and Environmental Impact Assessment

Future Perspectives

Bibliography

Glossary

Combustion refers to the overall process of initial “drying” (evaporation of water), “degasification” (pyrolysis), “gasification” (of the solid carbon), and “oxidation” of the combustible gases.

Fluidized bed refers to fluid dynamics in an intensively mixed system of solid particles and gas flow which keeps the particles in dynamic movement somehow comparable to “bubbling” or “circulating,” thus the mixture behaves much like a fluid. Fluidized bed is essentially a two phase flow phenomenon comprising a bulk of solids (e.g., quartz sand) which is kept in a fluidized state by an upward flow of gas (e.g., air) with velocities exceeding the minimum fluidization velocity. Thus the fluidized bed regime is characterized by intense movement of the solid particles resulting in high rates of heat and mass transfer. “Bubbling” and “circulating” fluidized bed systems can be distinguished depending on the gas velocity within the reaction chamber. In an “externally circulating” system the solids inventory is transported out of the reactor by gas flow at transport velocity.

Waste in the context of *waste-to-energy* typically refers to materials technically suitable for thermal treatment with ultimate oxidation of combustible matter. Treatment of waste may be necessary because of legal requirements for environmental protection (e.g., in order to destroy hazardous and other biologically or chemically reactive organic materials) or because of an economic advantage for recovery of energy by combustion of waste materials with a relevant calorific value (e.g., plastic wastes with high calorific value comparable to crude oil or bark and sawdust as a waste from processing wood). In advanced waste management regulations, the disposal of wastes exceeding 5% total organic carbon

(with some specific but limited exemptions) is legally prohibited in several European countries (e.g., in Austria, Germany, Switzerland).

Definition of the Subject and Its Importance

Waste is a complex issue. Disposal of waste in terms of “out of sight – out of mind” by burial on land or discharging into surface waters or by simple dumping onsite or in proximity of the origin of waste production has become a dramatic global problem and a cause for various forms of environmental degradation and severe damage to human health and the natural environment including fauna and flora. Increasing population and material wealth based on industrial production has led to tremendous waste problems. It has also led to greater understanding and the acknowledgment of resource limitations for an increasing human population on this planet. Therefore, the practice of dumping or disposal of waste must be viewed as wasteful, neglectful, and irresponsible to the needs of others and to those of future generations. It can also be viewed as a direct contradiction to “cultivating the earth” (the original concept for the human population on earth as described in the book of Genesis). Natural resources are often simply differentiated in “renewable” (based on the ongoing radiation of the sun onto the orbiting and rotating planet earth) and “nonrenewable” resources (particularly fossil resources such as mineral oil, natural gas, and coal). In any case, for both renewable and nonrenewable resources, whether material or energy, the issue of efficiency and recovery is vital for sustainability and the future quality of life on earth. The scientific basis for calculating resource efficiency, including recycling and recovery, is the first and second Law of Thermodynamics. Waste-to-energy is complementary to material recovery (including recycling). One unit of “ROE” (Raw Oil Equivalent) equals 40 MJ/kg or 40 GJ/t of calorific energy, which can be recovered by combustion and utilized in technical systems. Currently, about 90% of all presently consumed (nonrenewable) mineral oil in the form of various fuels goes directly into combustion. The use of mineral oil for production of materials such as plastic polymers allows for some recycling (recovery of materials for reuse in a recycling product) and also for ultimate recovery of the calorific value by waste-to-energy,

rather than its disposal in a landfill (or dumping elsewhere).

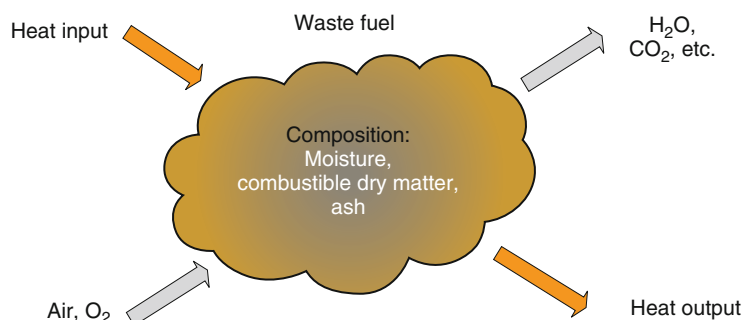
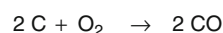
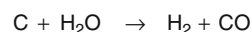
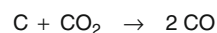
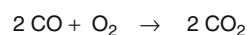
According to state-of-the-art, there are various technical systems for waste-to-energy. The type of system used depends on the specific type and mixtures of waste to be treated and the specific demand for various forms of energy onsite (e.g., heat at a given level of temperature for specific industrial production processes or for heating or cooling by operating an absorption process, demand for electricity). A continuous demand for cogeneration of power and heat is economically favorable for locating a waste-to-energy facility on such a site.

Fluidized bed is the most appropriate technology for a large variety of specific wastes (as well as for some type of biomass and low-grade coal). The basic principles of waste-to-energy and fluidized bed technologies, their applications, and limitations for waste treatment, some selected reference projects, technical aspects on emissions and environmental impact assessment, and future perspectives for further development in recovery of materials and energy from waste by fluidized bed technologies will be presented and discussed.

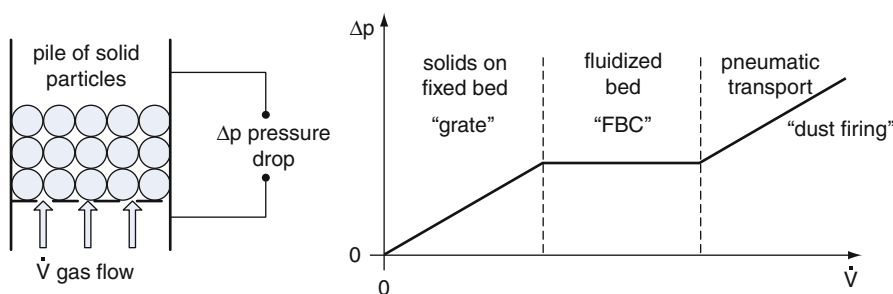
Introduction

The basic principle of waste-to-energy refers to the overall process of energy recovery based on energy conversion by incineration of waste materials. The overall process of heat production by incineration of waste (or any other solid fuel, such as wood or lignite coal) comprises the individual processes of initial “drying” (evaporation of water by heat input), “degasification” (pyrolysis – cracking of large organic molecules with subsequent release of volatile fractions), “gasification” (of the solid carbon by gaseous molecules such as oxygen, water, and carbon dioxide), and “oxidation” of the combustible gases (ultimately the oxidation of carbon monoxide and hydrogen). The terms “incineration,” “combustion” (like fuel in combustion engines), “thermal oxidation,” “thermal treatment,” and “gasification for generation of synthetic gas” (which will, of course, eventually be subject to some form of incineration) are often used interchangeably in waste-to-energy.

Figure 1 illustrates mass and heat transfer in the overall process of incineration with the four technically

Mass- and heat transfer**1 Drying****2 De-gasification (pyrolysis)****3 Gasification of carbon:****4 Oxidation of gases at high temperature:****Waste-to-Energy: Fluidized Bed Technology. Figure 1**

The four phases of the incineration process

**Waste-to-Energy: Fluidized Bed Technology. Figure 2**

Pressure drop in gas flow through static and fluidized beds and pneumatic transport

distinct phases of drying, degasification, gasification, and combustion. The overall process of incineration can be illustrated by frying/burning a steak: At first the meat will shrink due to loss of water (drying), secondly it will start to release odorous molecules and visible organic fumes due to continued frying until the remains of the steak gradually turn into black charcoal (completion of degasification), until finally only gray ash remains. After gasification of black carbon, only a small quantity of ash (containing phosphates, calcium, etc.) will remain from the original large piece of a juicy steak.

The process of combustion is based on heat and mass transfer which includes a supply of oxygen necessary for the oxidation as well as the removal of carbon dioxide as a major reaction product (which would

otherwise have the effect of a fire extinguisher). The necessary conditions for complete combustion are determined by the so-called 3T criteria, that is, "Temperature" (e.g., 850°C), "Time" (e.g., 2 s for the flue gas at the required minimum reaction temperature after the final supply of necessary oxygen) and "Turbulence" (rapid heat and mass transfer including thorough mixing of the combustible gases with the necessary oxygen in excess of the stoichiometric minimum need). The combustion process can effectively be monitored by the important parameters temperature, oxygen, and carbon monoxide.

The basic principles of fluidized bed reactors can be explained by the experiment (see Fig. 2) of a large number of particles (e.g., granular materials of defined particle size distribution) filled into a vertical cylinder

supported by a gas distributor at the bottom and a fluid (e.g., combustion air) passing upward through the bulk of solid particles. Initially, the particles remain at their original static position more or less densely packed. This state is like pieces of coal or wood lying on the grate of a furnace with air passing from underneath through the bed of solid fuel. This fluid regime is called “fixed bed.” With further increase of gas flow, the pressure drop increases up to the point when the bed with the particles starts “bubbling” and some particles will be lifted due to aerodynamic forces and start floating like a “fluid.” The fluidized bed behaves much like a liquid: It can be stirred; the bed surface remains horizontal when the vessel is tilted; the fluidized bed flows out like a liquid through a lateral opening; objects of greater specific gravity sink whereas those of lower specific gravity float; etc. [1].

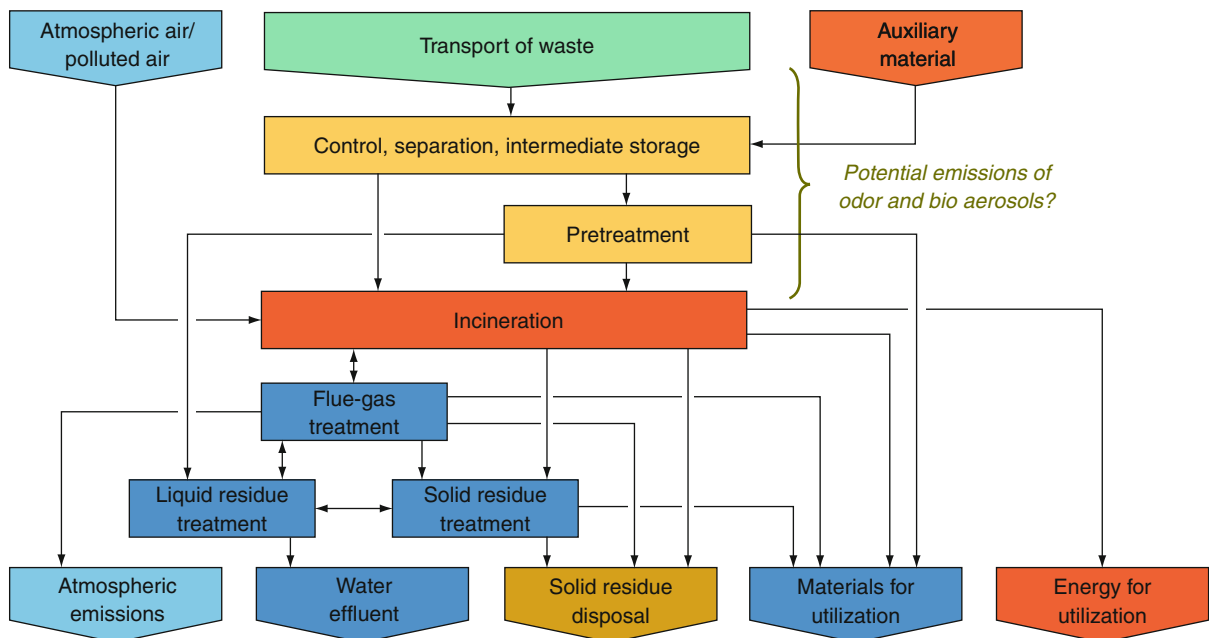
With even greater gas flow, the particle bed becomes more “fluidized” and various forms of “internal circulation” will occur. Eventually, further increase of gas flow will carry away the particles, like fine coal being pneumatically conveyed and fired in a gas flow.

The term “fluidized bed” obviously refers to fluid dynamics in an intensively mixed system of solid

particles and gas flow which keeps the particles in dynamic movement. Different systems can be described according to observation of the technical behavior as “bubbling,” “turbulent,” or “circulating” fluidized bed. Thus, various manufacturers refer to their system as a “bubbling bed,” a “turbo,” or a “circulating” fluidized bed. A fluidized bed system can be designed to induce specific forced circulation, for example, “internal circulation” (by specific geometry of the vessel – see example in Fig. 5) or “external circulation” (adjacent to the main combustion chamber is a large cyclone for separation of larger solid particles to be returned back into the fluidized bed and the hot flue gas leaving the cyclone through the afterburning zone). Technical examples are given in Figs. 4 and 10.

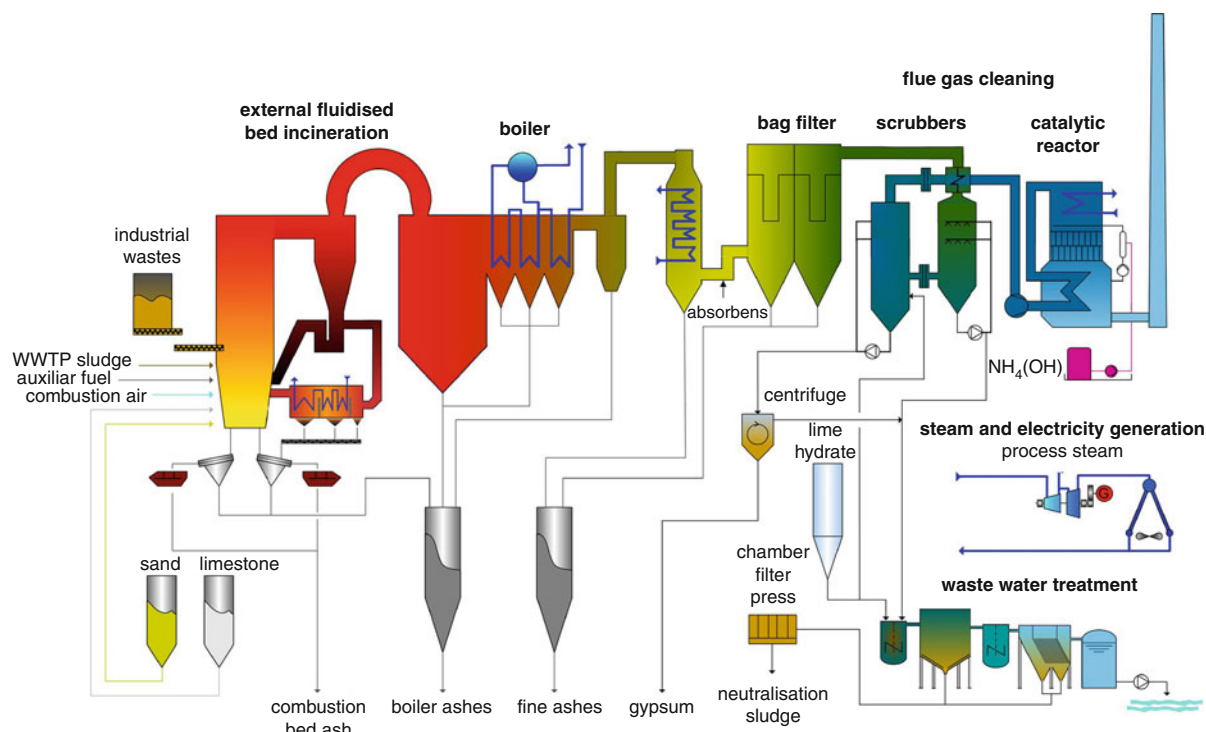
The complexity of large waste-to-energy projects is illustrated in Fig. 3. The incineration process (e.g., in the fluidized bed) may require some pretreatment of specific waste materials.

Figure 4 illustrates a waste-to-energy project with an “externally circulating fluidized bed combustion system.” The overall process design is also based on Best Available Technique (BAT) according to Austrian and EU requirements. Specific waste materials are delivered and stored



Waste-to-Energy: Fluidized Bed Technology. Figure 3

Overall scheme for the complexity of a waste-to-energy facility [2]



Waste-to-Energy: Fluidized Bed Technology. Figure 4

Example of waste-to-energy in an externally circulating fluidized bed system

in closed systems separately prior to incineration in the fluidized bed boiler. In this example the multistage flue-gas treatment process is effluent free because there is no ecologically acceptable discharge of chlorides into the river at the given location.

Technical Features and Limitations for Waste Treatment in FBC

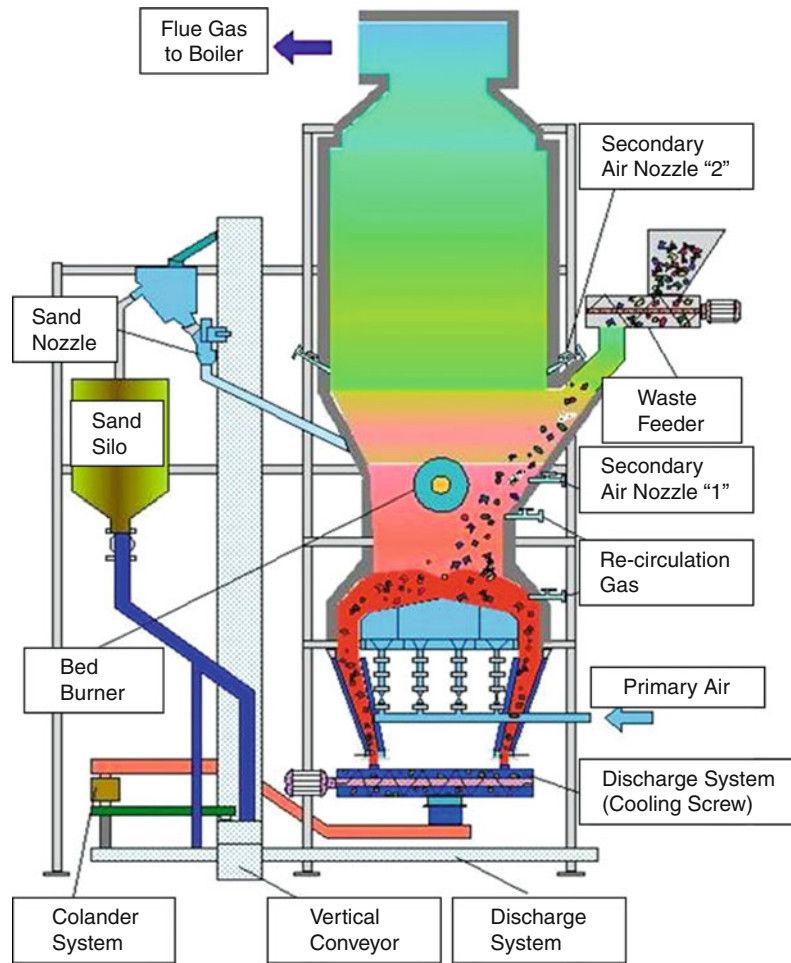
Fluidized bed systems allow for intensive mass and heat transfer between the solid particles to be incinerated and the flow of gas, thus enabling highest efficiency and complete chemical reactions. The three key parameters turbulence, temperature, and time can be technically controlled in fluidized bed systems within narrow ranges. Because of precise control of mass flow (air flows as well as fuel supply) the maximum temperature can be kept within a very narrow range of around 900°C, just above the minimum temperature necessary for complete gasification and combustion. Therefore, combustion in the fluidized bed systems allows for very

low formation of NO_x (nitrogen oxide) due to the absence of thermal formation by oxidation of nitrogen gas supplied with air for combustion. Such undesirable reactions do not occur at temperatures below 1,000°C.

The combustion process in a fluidized bed and the “freeboard” above it are complete. Thus, there are very low concentrations of unburned organic substances in the solid residues and in the flue gas.

Due to the precise process control, the excess air ratio (i.e., the excess oxygen content in the flue gas) after complete combustion can be kept very low including the option for controlled recirculation of cleaned flue gas extracted after the filter system (necessary removal of pollutants) back into the fluidized bed system. Reduced quantity of excess air minimizes the overall heat losses and atmospheric emissions of the system.

The following schemes in Figs. 5 and 6 illustrate the technical concept for precise control of gas flow in a waste-to-energy plant with an internally circulating fluidized bed as installed in Vienna: A small share of the



Waste-to-Energy: Fluidized Bed Technology. Figure 5
Example of an internally circulating fluidized bed system [9]

combustion air is the “primary air” necessary for fluidizing the bed material. It must be preheated in a heat exchanger in order to avoid condensation after mixing with high-pressure recirculation gas. The input of “secondary air” (for complete combustion of gas resulting from degasification and gasification of waste in the fluidized bed) can be varied by controlled utilization of recirculation gas and separately supplied ambient air. There is an additional supply of primary air (without need for preheating) by a separate ventilator for fostering internal circulation (“rotation”) of the fluidized bed.

The intensive turbulence also allows for extensive sorption processes at the given conditions in the fluidized bed system, especially of sulfur oxide being

chemically bound to calcium and magnesium oxides. These sorption agents are already present in waste material or fed into the fluidized bed, for example, in the form of calcium carbonate (limestone) or calcium hydroxide of defined fine particle size. This feature is also effectively used in the combustion of low-grade coal in fluidized bed systems with integrated reduction of sulfur dioxide emissions by about 90% (without secondary measures for sulfur oxide removal).

The precise control of fluidized bed systems also allows for a broad range of fuel properties with respect to calorific value. Fluidized bed systems can be designed and operated for average fuel mixtures ranging from about 5 to 8 MJ/kg up to values exceeding 40 MJ/kg (e.g., mixed plastic wastes). In extreme cases,

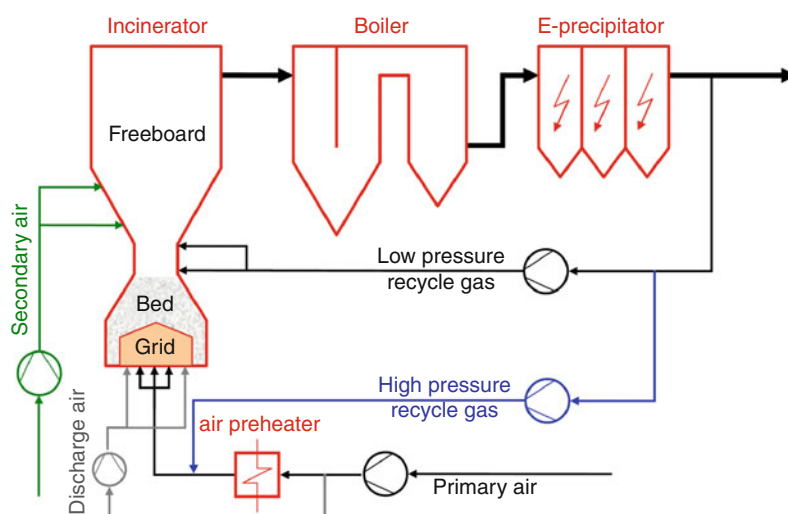
the average calorific value of waste to be incinerated can be below 3 MJ/kg (e.g., mechanically dewatered sewage sludge), with preheated combustion air (internal recovery of heat) and some waste flow being partly dried in a system utilizing low-pressure steam supplied from the own steam turbine.

The operating range also allows for reduced thermal load, e.g., as low as 50% of the maximum design capacity. The controllability of fluidized systems also allows for a very short shutdown time in case of an emergency because there is almost no unburned fuel within the fluidized bed system. In case of a temperature drop toward the minimum temperature requirement, auxiliary fuel supply or start-up burners can instantly be started. The emergency shutdown requires only a few

minutes (stopping of waste supply, reduction and shut-off of air supply).

One of the limitations of fluidized bed systems is the maximum allowable particle size suitable for controlled and safe mechanical feeding into the combustion system, for example, screening at 80 mm for limiting maximum size of pieces to be fed into the fluidized bed.

The maximum particle size of waste to be expected is usually the maximum size technically possible due to the collection system. Thus, in the case of some large containers in collection of municipal wastes rather large particle size must be expected (e.g., a large tire together with the wheel rim, mattresses with steel wires and springs, large containers with unknown



Waste-to-Energy: Fluidized Bed Technology. Figure 6

Example for air and gas flow in an internally circulating fluidized bed system [3]

Waste-to-Energy: Fluidized Bed Technology. Table 1 Comparison of technologies for waste-to-energy

Parameter	Incineration technology		
	Grate	Fluidized bed	Rotary kiln
Maximum fuel capacity per line	Approx. 90 MW	Approx. 160 MW	Approx. 40 MW
Excess air (oxygen) ratio (specific quantity of flue gas)	Medium	Low	High
Range of calorific value for waste and waste-derived fuel	Low	High	Medium
Fuel processing requirements	Low	High	Medium
Controllability of incineration and shutdown operation	Medium	High	Low

Waste-to-Energy: Fluidized Bed Technology. Table 2 Allocation of selected wastes to specific incineration technologies

Type of waste	Incineration technology		
	Grate	Fluidized bed	Rotary kiln
Mixed municipal solid waste (Residual mixed waste)	Well-suited	Pretreatment required	Suitable
Sewage sludge	Limited in terms of quantity	Well-suited	Suitable
Screening material from sewage treatment	Suitable	Pretreatment required	Limited suitability
Shredded plastic wastes	Limited in terms of quantity	Well-suited	Limited suitability
Scrap tires (from vehicles)	Limited suitability	Unsuitable	Limited suitability
Shredder waste (from metal scrap recycling)	Limited in terms of quantity	Well-suited	Limited suitability
Shredded waste wood (e.g., from processing bulky wastes)	Well-suited	Well-suited	Suitable
Lacquer and paint sludge,oil sludges	Unsuitable	Suitable	Suitable
Wastes in closed containers (e.g., medical and laboratory waste)	Limited suitability	Unsuitable	Suitable

**Waste-to-Energy: Fluidized Bed Technology. Figure 7**

Waste-to-energy facilities at Simmering, Vienna: 4 fluidized bed and 2 rotary kilns

content, etc.). Therefore, this type of waste with broad ranges of size and properties might require presorting or mechanical shredding prior to combustion in a specific fluidized bed system. The specific limitations

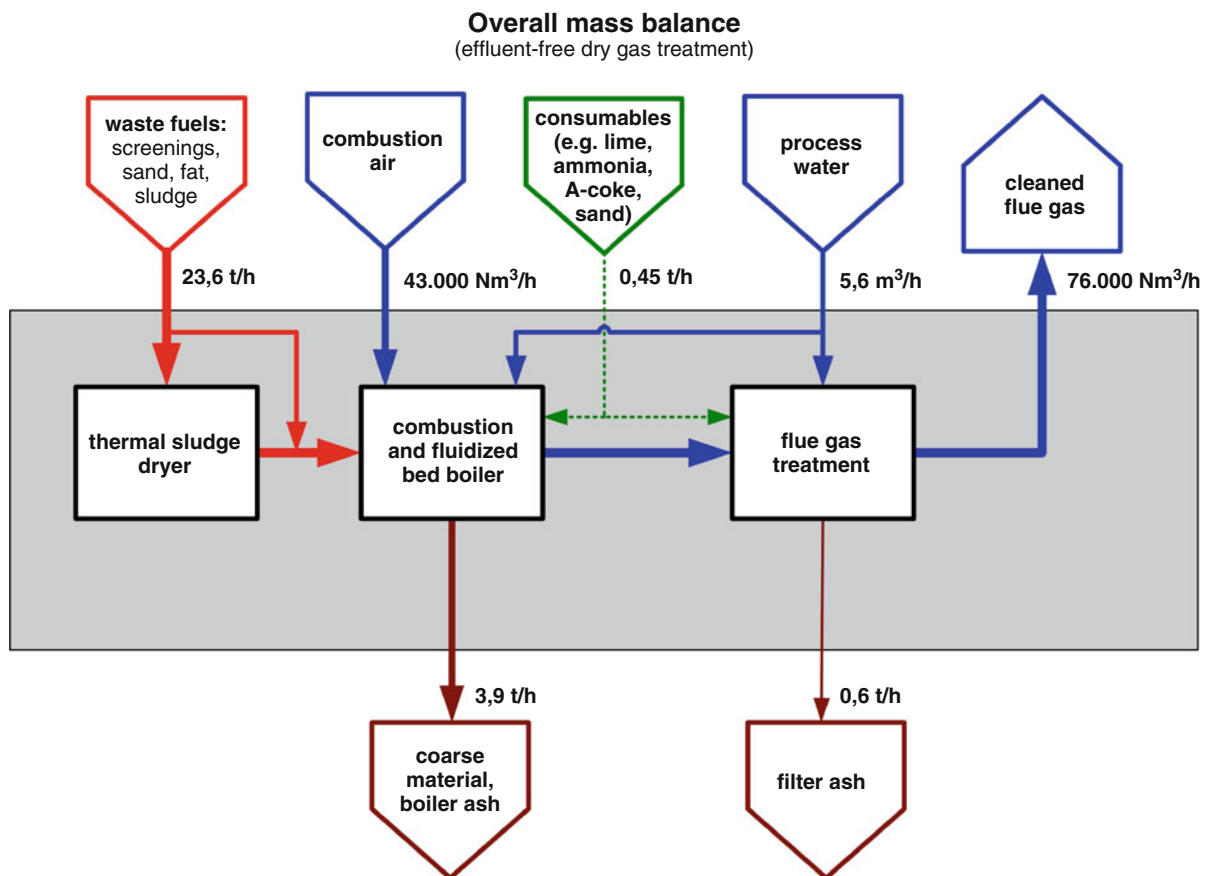
on the maximum size and shape of waste for the technical feeding system as well as for the removal of coarse inert parts after incineration (which is removal of bed material and ashes) must be considered. In comparison

to fluidized bed systems, the typical grate systems for waste incineration are more suitable for typical large parts in unsorted municipal solid waste.

Some waste materials are neither suitable for fluidized bed nor for grate systems due to a broad range of (often unknown) potentially hazardous properties. Therefore, large rotary kilns with an integrated afterburning chamber are typically used for various mixes of hazardous wastes or closed containers with partly unknown content (e.g., infectious and other medical wastes, laboratory wastes). Tables 1 and 2 provide a general comparison of fluidized bed systems with other important systems for thermal waste treatment and waste-to-energy such as grate systems and rotary kiln (including an afterburning chamber).

Fluidized bed systems operate with high turbulence in the combustion process. This allows for the recovery

of larger size of clean (like sandblasted) solid particles from the coarse bed material, but will also yield large quantities of very fine ash (although with a low content of unburned residues). This is of advantage with respect to maximum recovery of specific materials, but it may be a general disadvantage for disposal of untreated solid residues at lower costs. The quantity of (naturally fine) fly ash in a grate system firing typical municipal waste may be only about 10% of the total ash content, whereas in case of a fluidized bed, fly ash may be about 60–80% of the total ash quantity. Therefore, some fluidized bed systems use additional cyclones for particulate removal from flue gas in the boiler system at about 400°C in order to recover most of the (coarse) fly ash in a nonhazardous quality (no condensed pollutants such as metal chlorides at such temperature and no water-soluble chemical reaction products before



Waste-to-Energy: Fluidized Bed Technology. Figure 8

Overall mass balance for a fluidized bed system for residues from sewage treatment

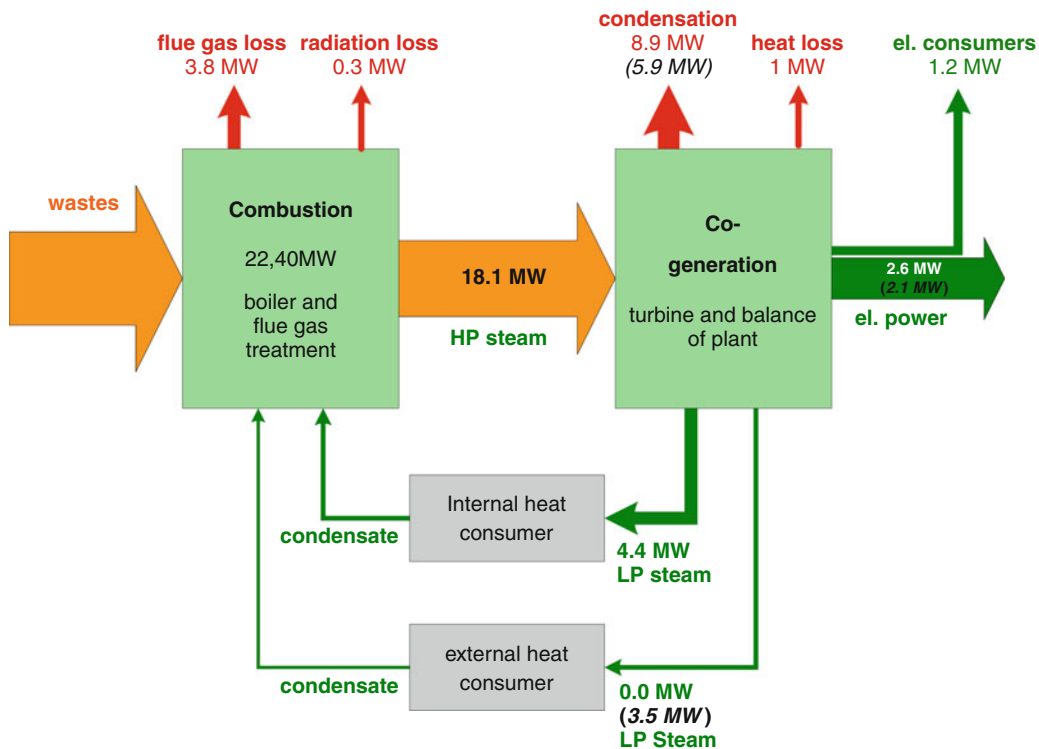
injection of sorption agents into the flue gas). Slag from municipal waste combustion on grate systems is generally qualified as a nonhazardous material with the option of disposal at low cost in landfills.

Specific attention must be given to the potentially high content of metallic aluminum in waste combustion for two reasons: economic opportunity for recovery and recycling of aluminum and the potential risk of operational problems and hazardous properties of fine ash. Aluminum content in municipal waste has significantly risen throughout the last years (packaging, laminated materials, various aluminum products). Aluminum alloys have a melting point around or somewhat below 700°C; thus, metallic aluminum will be completely “dissolved” at such temperatures in the turbulence of fluidized bed systems. The very fine aluminum droplets (with a very stable surface of a thin layer of protective aluminum oxide) will be carried by the flue gas and will eventually either adhere in layers on surfaces with lower temperatures (also causing risk for sudden exothermic reactions) or become part of the

removed fine ashes. These ashes have naturally high alkalinity and thus the fine aluminum will react in case of contact with water with the release of hydrogen gas (up to 1.3 l of H₂ per g of Al). Therefore, systems with temperatures below 650°C in the fluidized bed allow for recovery of aluminum parts in removal from coarse bed material. Systems operating at higher temperatures must strictly limit the maximum allowable content of metallic aluminum in ashes by operating with appropriate waste fuel mixes (including automated removal of nonferrous metallic parts before entering a combustion process).

The types and installed capacities of specific waste-to-energy technologies may vary significantly between various regions. According to many years of experience in Austria and some other European countries, the different technologies for thermal waste treatment and waste-to-energy are needed and applied.

The example of the city of Vienna and the more industrialized state of Upper Austria illustrate the successful operation of various technologies for



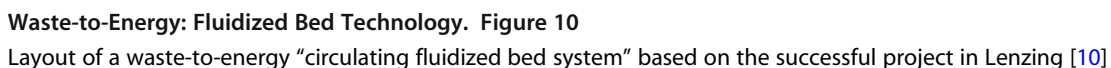
Waste-to-Energy: Fluidized Bed Technology. Figure 9

Energy balance for a fluidized bed system for residues from sewage treatment

and other industrialized countries [1]. Since the late 1960s and 1970s, treatment of waste water and sewage has become a high priority for municipalities and large industry. Subsequently, the need for sewage sludge treatment became obvious. Thus, an increasing number of fluidized bed systems were installed throughout the Western countries in order to treat those residues by thermal conversion (incineration). The first significant price increase for fossil fuels encouraged industry to utilize their burnable wastes (e.g., bark from processing wood in the pulp and paper industry) together with residues from mandatory wastewater treatment in appropriate boilers. Of course, these waste-to-energy plants also needed to be suitable for burning various types of coal for heat and power supply to the industrial production.

In 1980, two fluidized bed incinerators, based on the principle of "bubbling bed" (stationary fluidized bed), were installed in Vienna for treatment of mechanically dewatered sewage sludge with auxiliary fuel of waste oil or heating oil as needed. These

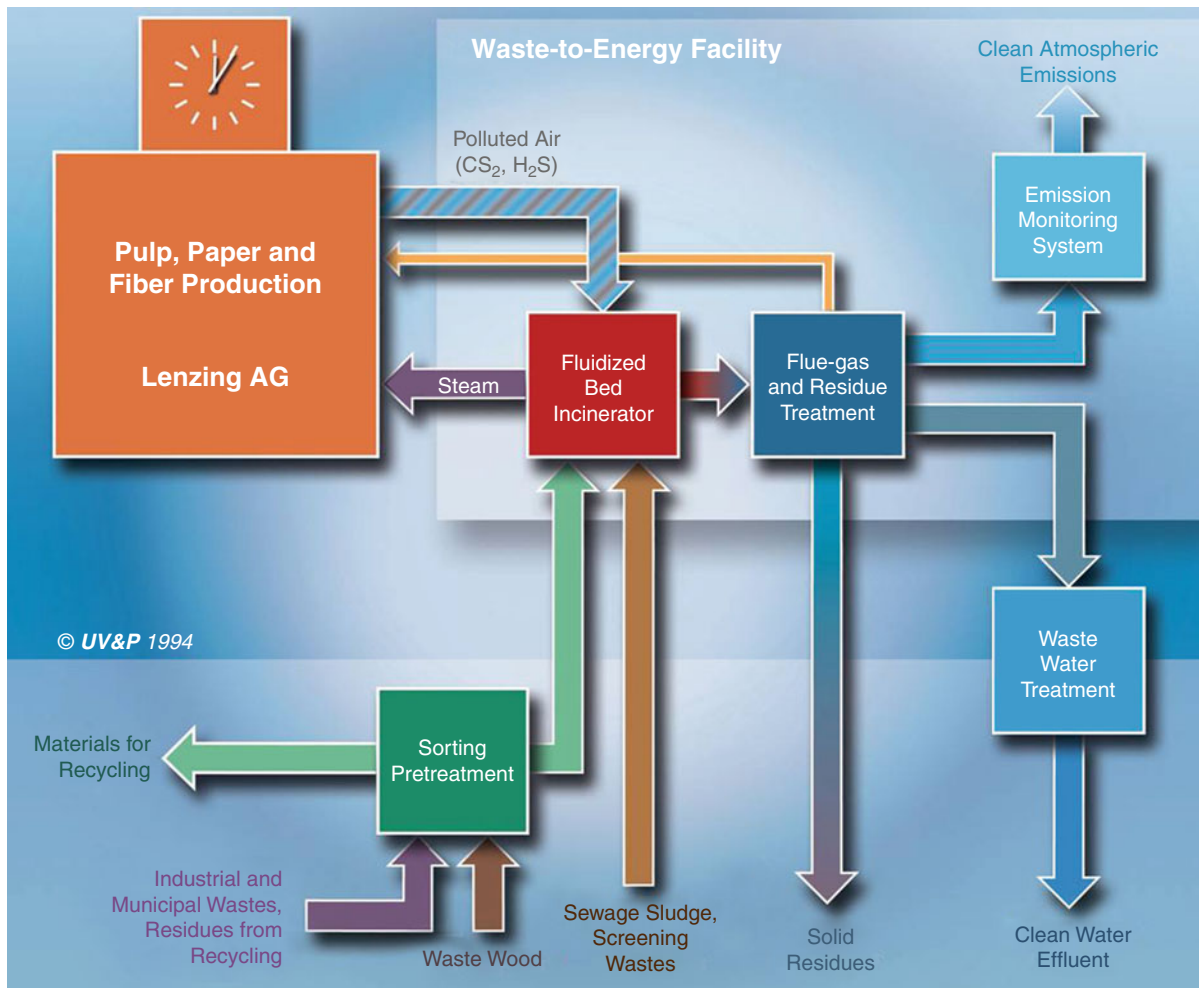
Fluidized bed systems have been used for gasification and combustion for many decades throughout Europe



waste-to-energy plants are still in successful operation today, together with an additional similar fluidized bed system installed in 1992 [4]. For the purpose of standby and guaranteed treatment of all sludge generated at all times and for the purpose of waste-to-energy from pretreated municipal wastes an additional fluidized bed incinerator was installed at the same site in 2003. This incinerator is also a stationary system, but with a “rotating bed,” also referred to as an “internally circulating fluidized bed” system (see Figs. 5 and 6). The overall performance of the four fluidized bed incinerators together with the two rotary kilns for hazardous wastes (plus a small rotary system for pretreatment

of consumer batteries) at the site of Simmering is presented in Fig. 7. It should be noted that these facilities are situated next to the sewage treatment plant in the central area for vegetable production in Vienna – with full public acceptance of local residents.

The overall mass and energy balance for thermal treatment of residues from a large municipal wastewater treatment in a fluidized bed system is presented as example in Figs. 8 and 9. In order to operate the plant solely on residues from municipal sewage treatment (without auxiliary fuel), part of the mechanically dewatered sludge is dried in a machine operated with low-pressure steam from the waste-to-energy cogeneration facility.



Waste-to-Energy: Fluidized Bed Technology. Figure 11

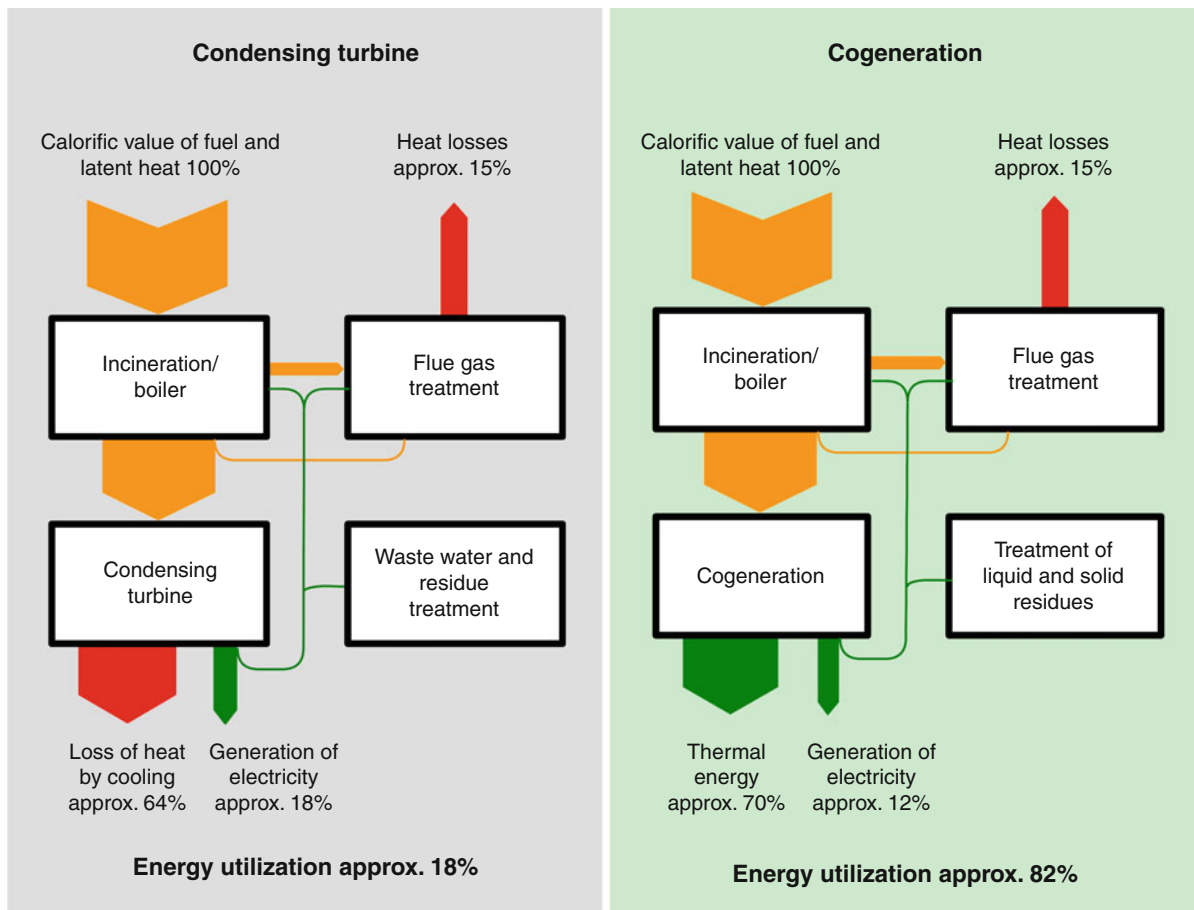
Scheme for integrated waste-to-energy in the industrial production at Lenzing



Waste-to-Energy: Fluidized Bed Technology. Figure 12
View of the industrial site Lenzing within a region of tourism and organic farming [2]

Fluidized bed systems for waste treatment and recovery are successfully in operation with small fuel (i.e., thermal generation) capacities below 5–10 MW in Europe. Examples include energy recovery with ORC (Organic Rankine Cycle) and simultaneous recovery of the mineral sand fraction from bituminous waste products at an industrial production site as well as a small plant for incineration of various solid, pasteous, and liquid hazardous wastes.

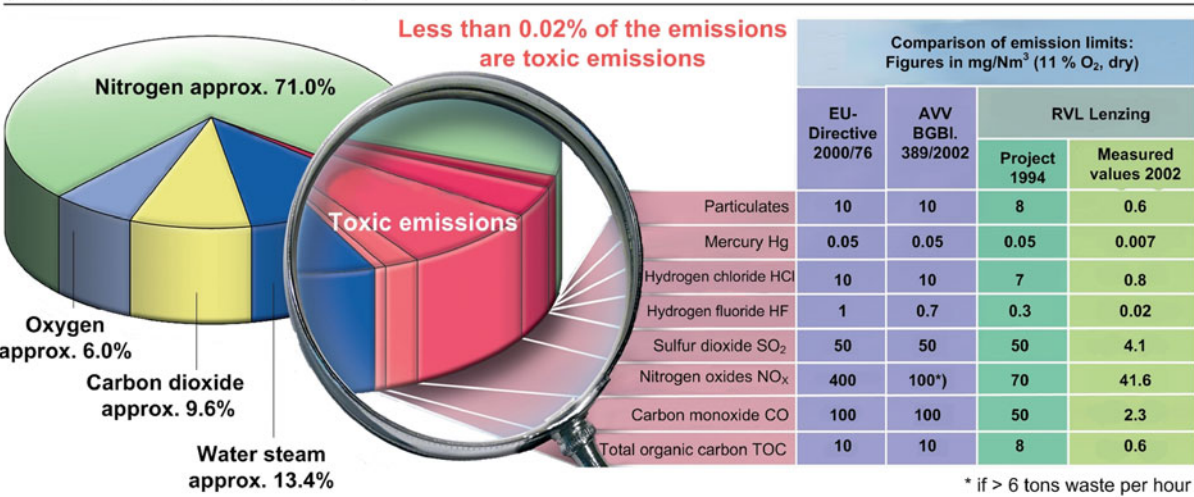
Large waste-to-energy facilities (up to 160 MW fuel capacity per line) for cogeneration of process heat and electricity at industrial sites are typically based on “externally circulating fluidized bed systems,” as illustrated by the example of Lenzing in Upper Austria (see Figs. 11 and 12). This project was developed and planned in 1993 with subsequent heated



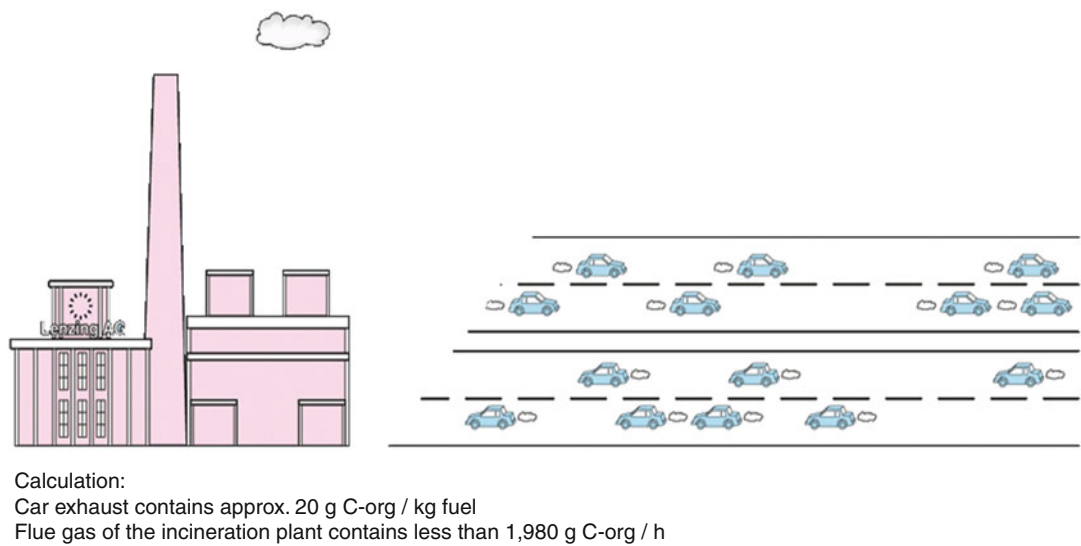
Waste-to-Energy: Fluidized Bed Technology. Figure 13

Energy utilization with a condensing turbine compared to cogeneration presented by the example of waste-to-energy at the more appropriate industrial site Lenzing

Control of cleaned flue gas from waste incineration
(Example: RVL Lenzing)



Waste-to-Energy: Fluidized Bed Technology. Figure 14
Composition of atmospheric emissions and standards for the waste-to-energy project in 1994 in comparison to actual performance in 2002 and strictest legal emission standards

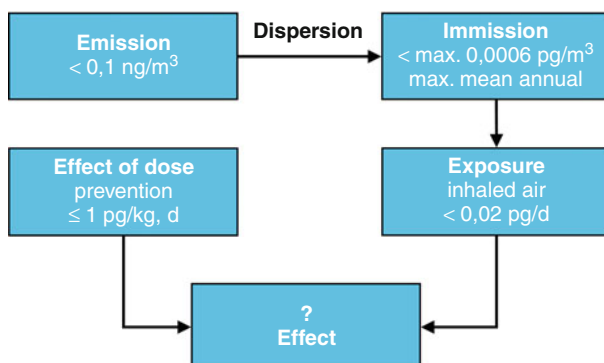


Waste-to-Energy: Fluidized Bed Technology. Figure 15
Illustration for organic pollutants from the waste-to-energy plant compared with passenger cars (example [2])

discussions in the media and public hearings, nevertheless resulting in a successful start-up of operation in 1998. Based on the experience and know-how derived during optimization of this first plant of its kind, several more such large waste-to-energy plants have been built or are currently under construction by the supplier, ANDRITZ Energy & Environment (see Fig. 10).

$$1 \text{ ng} = 10^{-9} \text{ g} = 0,000\,000\,001 \text{ g}$$

$$1 \text{ pg} = 10^{-12} \text{ g} = 0,000\,000\,000\,001 \text{ g}$$

**Risk assessment:**

The precautionary protection standard (federal health bureau, Berlin) is 1 pg / kg weight and day (factor 1000 safety compared with NOEL - No Observable Effect Level). In the worst case a person inhales 0,012 pg dioxin with ca. 20m³ air per day. In comparison the protection standard for a 60 kg person is 60 pg/day. Therefore the inhaled maximum is less than $\frac{1}{1000}$ of the protection standard. Thus the additional risk can be considered irrelevant.

No increased health risk due to very small concentration of dioxins in the cleaned flue-gas of the waste incineration facility.

Waste-to-Energy: Fluidized Bed Technology. Figure 16

Risk assessment for atmospheric emissions of "dioxins" (example [2])

Assumption: 3 cigarettes per day = marginal value of impact (no effect observable)
Thus the precaution protection standard ($\approx 1/1000$) is equivalent to 1 cigarette per year.



Conclusion : An incineration plant with multi-stage flue gas cleaning is factor 1.000 below the protection standard
thus = equivalent to 1 additional cigarette in 1.000 years

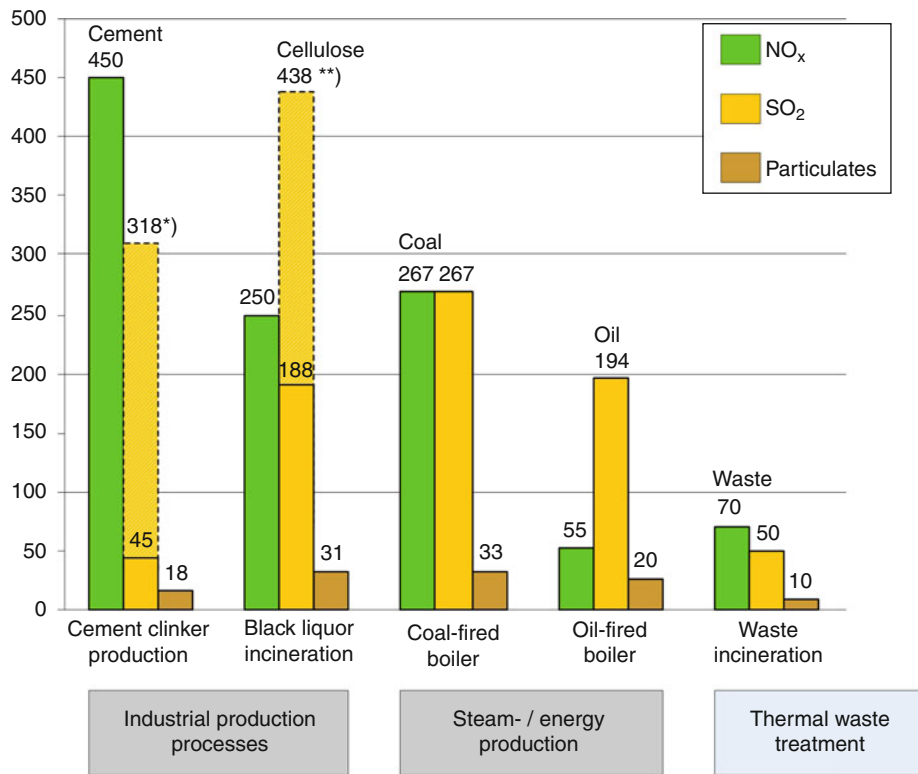
Waste-to-Energy: Fluidized Bed Technology. Figure 17

Illustration for risk assessment in comparison with cigarettes

Waste-to-Energy: Fluidized Bed Technology. Table 3 Development of emissions from waste incineration according to state-of-the-art in Austria and Switzerland [2] (figures in $\text{mg}/\text{m}_\text{N}^3$, PCDD/F in (TEQ) $\text{ng}/\text{m}_\text{N}^3$)

Year	Dust	Cd	HCl	SO ₂	NO _x	Hg	PCDD/F
1970	100	0.2	1,000	500	300	0.5	50
1980	50	1.1	100	100	300	0.2	20
1990	1	0.005	5	20	100	0.01	0.05
2000	1	0.001	1	5	40	0.005	0.05

Figures for plants > 50 MW in mg / m³_N at 11 % O₂, dry.



*) Special legal permit available up to 318 mg / m³

**) Depends on the process

Waste-to-Energy: Fluidized Bed Technology. Figure 18

New source performance standards for flue gas from different industrial incineration processes in Austria

Emissions and Environmental Impact Assessment

Emissions and the assessment of environmental impact from a large waste-to-energy plant with fluidized bed technology can be presented based on the extensive public discussions in 1993/94 in Austria around the project in Lenzing. There were fundamental objections to such projects because of the fear of pollution from waste incineration by the local population and also due to strong opposition from the competing waste disposal industry outside of the planned project. Furthermore, the legal ban on disposal of wastes exceeding 5% TOC (total organic carbon) in landfills had not yet been established and was open for controversial public discussion.

Figure 11 summarizes the basic concept and technical features of the waste-to-energy project based on

a fluidized bed system in Lenzing for comprehensive public information.

In extensive public discussions, justification for the considerable size of the project (thermal power of approx. 110 MW which allowed for the incineration of up to 1,000 t of waste per day) and the choice of the location (surrounded by organic farms and by the tourist region of Lake Attersee as depicted in Fig. 12) was successfully presented using the following three arguments:

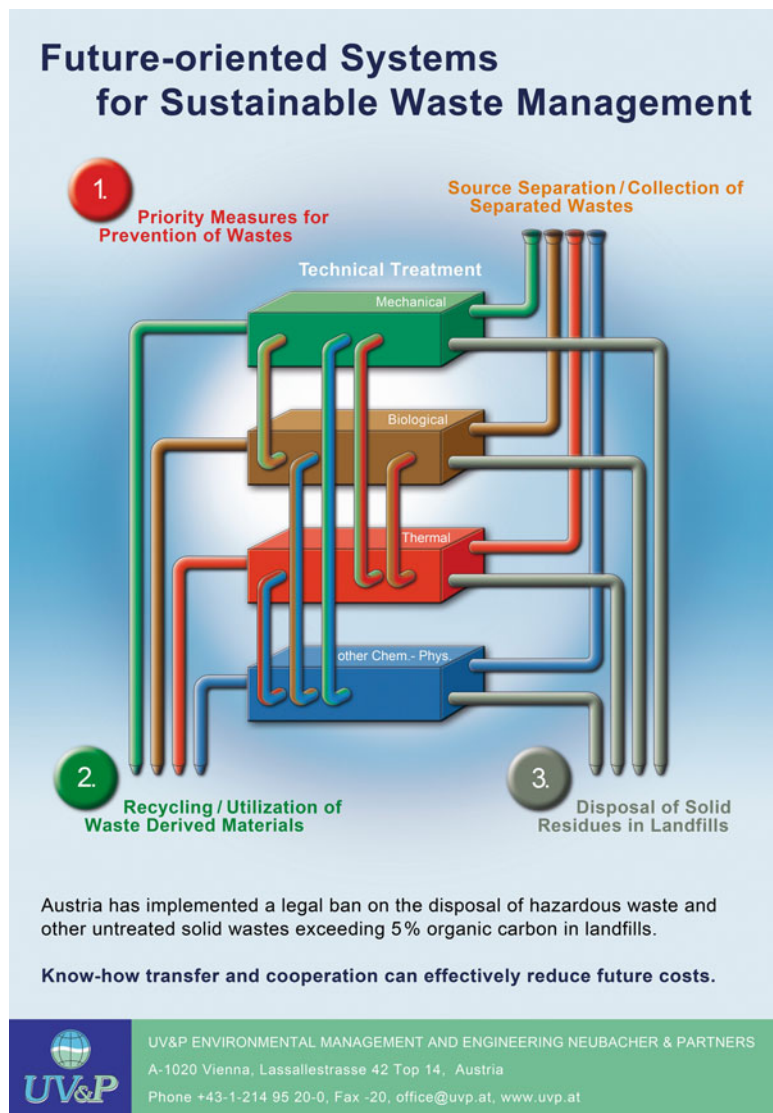
1. Steam requirement for the industrial production of Lenzing (elimination of two old boilers from 1938/1939 at the site of the new plant, substitution of imported natural gas by waste-to-energy)

2. Improvement of local air quality by utilization of the malodorous air emitted from viscose fiber production for the necessary combustion air in the waste-to-energy boiler
3. Avoiding future disposal of reactive organic wastes in landfills (so-called reactor dumps), thus avoiding uncontrollable emissions from landfills and contaminated sites for future generations

The argument in favor of the proposed capacity at the highest level of energy efficiency in a fluidized

bed boiler with high steam parameters (80 bar, 500°C) for cogeneration (based on continuous demand for process steam throughout the year) was illustrated by the comparison to a municipal waste incineration plant (with maximum production of electricity in a condensing turbine) as indicated in Fig. 13.

The fear of uncontrolled atmospheric emissions and possible adverse health impacts was of highest concern to local citizens in public discussions. Therefore, providing continuous measurements and monitoring of all relevant parameters was a priority. The development



Waste-to-Energy: Fluidized Bed Technology. Figure 19
 Integrated system for sustainable waste management

Specific treatment and utilization of:

- separate collection and treatment of infectious and other hazardous organic wastes (hazardous waste incineration)
- mixed municipal waste (garbage)
- bulky municipal wastes
- construction and demolition wastes
- materials for recycling (e.g. paper, cardboard, glass, PET)
- green wastes
- food and kitchen wastes
- materials for specific treatment (e.g. batteries, tires)

**Waste-to-Energy: Fluidized Bed Technology. Figure 20**

Separate collection of municipal wastes for recovery of materials and energy

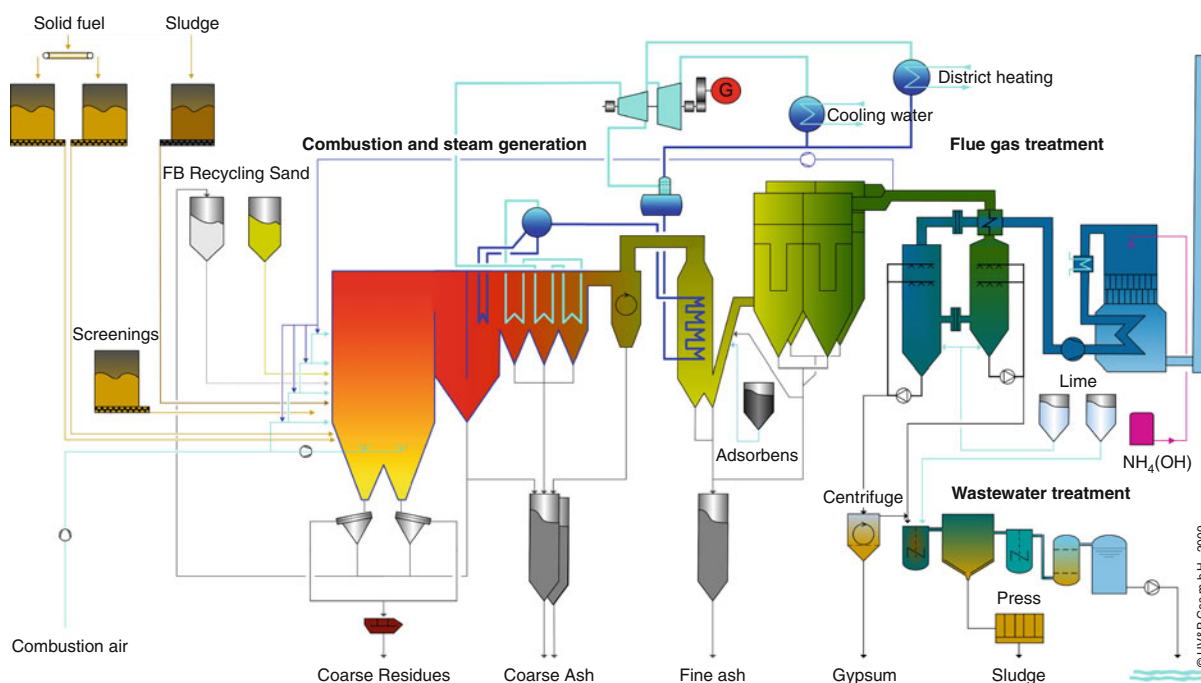


Waste-to-Energy: Fluidized Bed Technology. Figure 21
Projected aerial view of waste-to-energy plant in Austria
(design by architect H. Ganguly)

and planning of the waste-to-energy project was based on the most stringent emission standards worldwide (!). The public was also informed about the composition of the cleaned flue gas leaving the (already existing 152 m high) smokestack (see also Fig. 14).

Subsequently, public concerns were expressed regarding the unknown organic pollutants as indicated by the summary parameter TOC (total organic carbon). However, the comparative calculation shows that the maximum allowable emission through the smokestack is equivalent to the organic carbon emissions of 14 passenger cars driving somewhere in the region (e.g., on the nearby highway A1, see Fig. 15).

The public discussion also showed great concern of some citizens for the need to analyze each delivery of incoming waste in order to be informed of the materials to be incinerated. However, the public has also been informed regarding the mass and emission balances for all of the six critical elements for thermal waste treatment processes. These include chlorine (predominant pollutant in various waste), fluorine, sulfur (predominant pollutant in some wastes, high load in the off-gas used for combustion air in the case of Lenzing), mercury (extremely volatile and toxic), cadmium (medium volatile and very toxic), and chromium (not volatile, but toxic in the water-soluble form of Chromium^(VI)). Thus, further chemical analyses are to be carried out only in the case of unknown production wastes and if there is suspicion of potentially high pollution load. The planned waste-to-energy plant will be equipped and operated to comply with the emission standards under all circumstances, also with the understanding that waste incineration in a fluidized bed can be halted within minutes in case of an emergency.



Waste-to-Energy: Fluidized Bed Technology. Figure 22

Waste-to-energy plant for processing municipal solid waste and residues from recycling and treatment of sewage in the region of Linz

Finally, some citizens (including lobbyists opposed to the waste-to-energy project) articulated concerns about the most dangerous emissions in the form of “dioxins.” The answer was already prepared by a risk assessment based on a maximum allowable emission of 0.1 ng TE/m^3 (TE = toxicity equivalent of poly-chlorinated di-benzo-dioxins and -furans expressed in 2,3,7,8 TCDD) and on the results of a dispersion model with meteorological data for an entire year measured onsite. The overall response presented by independent medical experts indicated a hypothetical, but extremely small additional risk (see Fig. 16). This risk was illustrated to the public by comparing the numbers with smoking a cigarette. Comparatively speaking, the amount of dioxin emitted from the waste-to-energy plant equals smoking one additional cigarette in 1,000 years, if the precautionary value for protection of health is supposed to be the maximum of one cigarette per year (see Fig. 17).

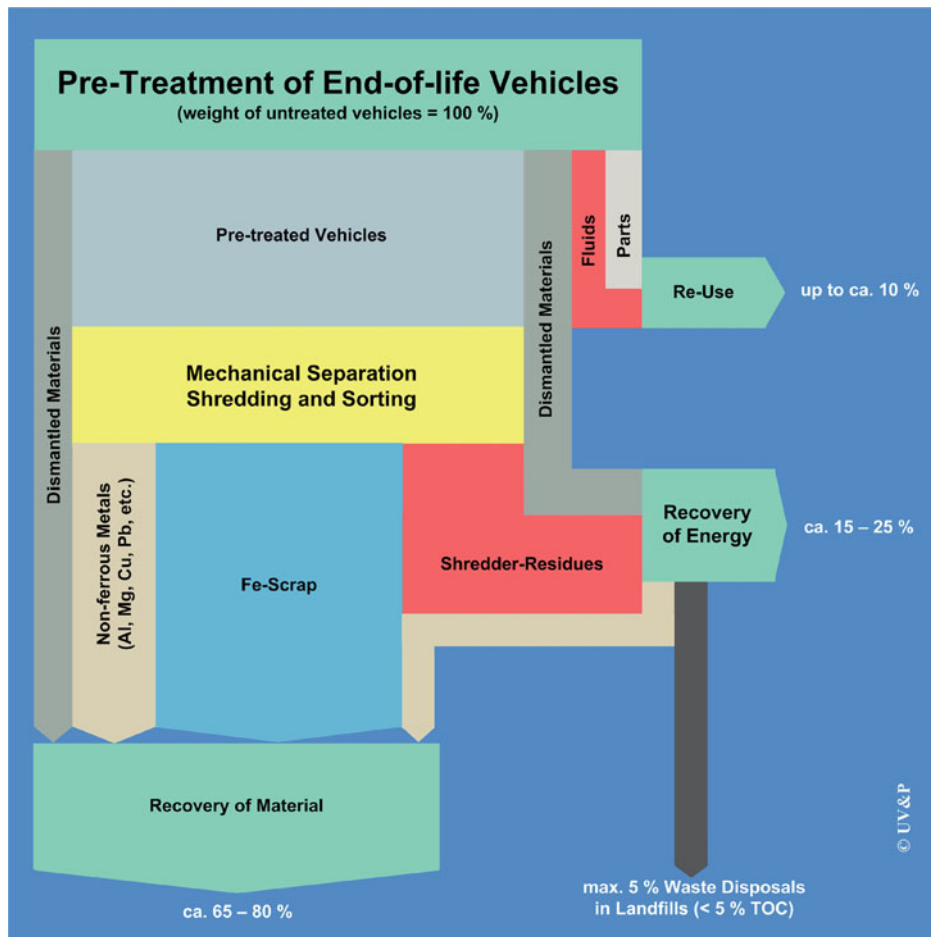
The discharge of dissolved chlorides within the small quantity of treated wastewater effluent (after removal of all heavy metals in multistage

wastewater treatment) has been of no public concern due to the relatively small quantity in comparison to chlorides used for ion-exchange in the softening of hard water and salt applications on roads against ice in winter.

The negative public opinion regarding emissions from waste incineration is generally due to several factors: ignorance of the facts and figures and past achievements in making waste-to-energy an environmentally friendly option (as indicated by the development of atmospheric emissions according to state-of-the-art in Austria and Switzerland – see Table 3). The comparison of strict emission standards for various new sources of industrial production and heat and power generation indicates the outstanding environmental standard of waste-to-energy (Fig. 18).

Future Perspectives

Sustainable waste management will require an integrated approach for recycling and recovery based on material separation. Various specific technologies for



Waste-to-Energy: Fluidized Bed Technology. Figure 23

Processing of end-of-life vehicles in Austria and EU that includes material and energy recovery from automobile shredder residues

treatment of separated wastes are needed including thermal processes in context with recovery of materials and energy as illustrated in Fig. 19.

In the case of municipal waste management, waste-to-energy is needed for residues from recycling and material recovery as well as for mixed municipal wastes (Fig. 20). For practical purposes, fluidized bed systems for waste-to-energy at appropriate locations (with continuous demand for heat throughout the year, industrial infrastructure, and competence in operation and maintenance) should be considered as an option for integrated sustainable waste management in the future. The latest project in Central Europe for industrial waste-to-energy has been planned for integration into an existing cardboard recycling and production facility with a continuous

demand of steam and electricity throughout the year. The planned project will substitute for imported natural gas and will use reject materials and other residues from paper and cardboard recycling for waste-to-energy in a fluidized bed boiler. The planned thermal capacity is 160 MW suitable for up to 450,000 t of waste fuel per year. The architectural design of the planned facility illustrates the option for aesthetic harmony with the surrounding environment (see Fig. 21).

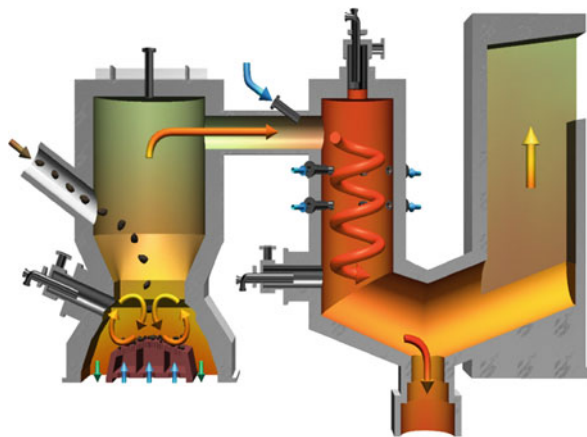
The following example of waste-to-energy for district heating and power generation illustrates integrated waste management for the region of Linz in Upper Austria (see Fig. 22). This plant will start operations in 2011 [5]. It is designed for a fuel capacity of 72.6 MW and allows for the utilization of 200,000 t of various residues from

recycling and sewage treatment as well as for mechanically pretreated mixed municipal solid waste. The fluidized bed system comprises a sophisticated air control and distribution system that allows for highest efficiency in combustion and also for very low emissions [6]. This plant is located in the central area of the city of Linz and has been permitted based on a positive

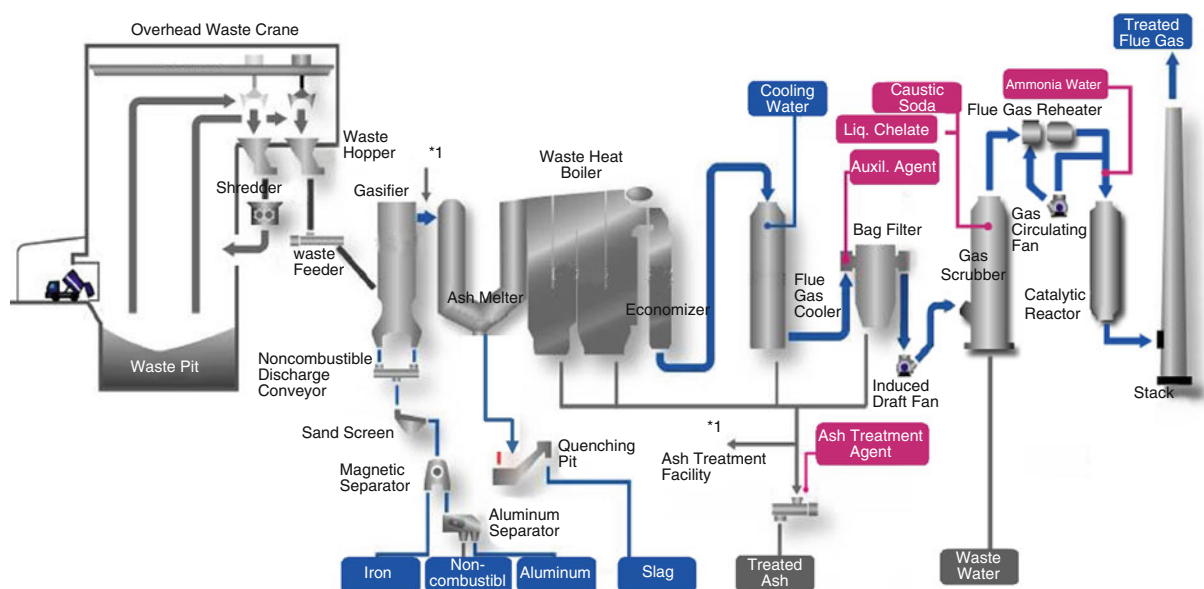
Environmental Impact Assessment without any objection from neighbors and citizens.

The future challenges for waste-to-energy in fluidized bed systems include recovery of valuable materials prior, during, or after the thermal treatment process. This will be especially relevant for the treatment of shredder residues from end-of-life vehicles and other wastes containing metals and rare elements. The increase of recovery mandated in the European Directive 2000/53/EC on end-of-life vehicles to 95% (including 85% of reuse and material recovery) by 2015 will be a significant challenge for adaptation and further development of fluidized bed technologies (see Fig. 23). Such a solution may be provided by the Ebara internally circulating fluidized bed system (operating at around 600°C) followed by high-temperature combustion at around 1,400°C in an ash melting furnace that has been developed and is operating in Japan [7] (Figs. 24 and 25).

Waste-to-energy plants require a continuous fuel supply. However, waste-to-energy plants need planned shutdown times of typically 3 weeks or longer for maintenance, plus some for unexpected necessary repairs. Therefore, it is necessary to provide for intermediate storage of the feedstock to the plant.

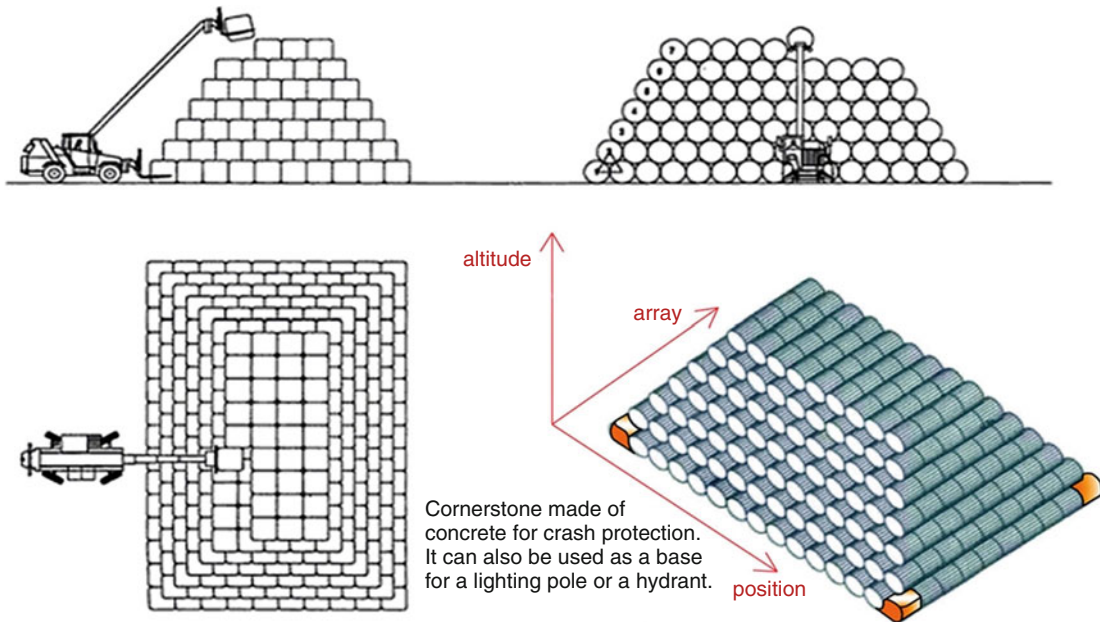


Waste-to-Energy: Fluidized Bed Technology. Figure 24 TwinRec core components gasifier and ash melting furnace [5] (© Ebara Environmental Plant Co., Ltd.)



Waste-to-Energy: Fluidized Bed Technology. Figure 25

Combined energy and material recovery at the Aomori plant [5] (© Ebara Environmental Plant Co., Ltd.)



Waste-to-Energy: Fluidized Bed Technology. Figure 26

Logistic and intermediate storage in view of environmental protection and fire safety (Patent applications A 1037/2008, PCT/EP 2009/050238)

Technical guidelines have been developed for environmentally safe storage of different high-calorific and organic wastes [8]. As discussed in other entries in the waste-to-energy section of this encyclopedia, the bunker of a waste-to-energy plant can be designed to store up to 2 weeks of incoming wastes; when longer periods of storage are required, the wastes can be baled as illustrated in Fig. 26, based on proven know-how and experience in Europe. Cylindrical bales of approximately 1.2 m diameter and height can allow for a storage capacity of up to 60,000 t of solid wastes per hectare. Specific features for prevention of fire hazards and adverse environmental impacts of this method of storage and transport of wastes have been developed in Austria [8].

Bibliography

Primary Literature

1. Werther J (1992) Fluidized bed reactors. In: Ullmann's encyclopedia of industrial chemistry, vol B4. VCH, Weinheim
2. White Book Waste-to-Energy in Austria: Figures, Data, Facts (2010) Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management, Vienna
3. Strauss T et al. (2005) Start-up and operation optimization of a 39 MW_{th} bubbling fluidized bed incinerator for domestic waste and sewage sludge. Fernwaerme Wien, Vienna
4. Stubenvoll J et al. (2002) Stand der Technik bei Abfallverbrennungsanlagen, Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management, Vienna
5. Zschetzsch A (2011) RHKW Linz: engineering, contract specifications, supervision of erection and commissioning, AZ² Technisches Büro, Linz in cooperation with UV&P, Vienna
6. Seemann B (2010) BioCOM® – fluidized bed combustion technology for thermal utilisation of biomass and waste. HS energieanlagen, Freising
7. Seliger A, Steiner C (2004) Waste gasification in practice: TwinRec fluidized bed gasification and ash melting – review of four years of plant operation. In: IT3'04 conference, Phoenix, 10–14 May 2004
8. Huber KH et al. (2007) Anforderungen an die Zwischenlagerung von heizwertreichen Abfaellen, Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management, Vienna
9. Integral (2004) The fluidized bed incinerator #4 for the treatment of municipal waste and sewage sludge. INTEGRAL, Achau
10. Hashemian M (2011) ECOFLUID RC and POWER FLUID® RC fluidized bed technology for "Energy from Waste" ANDRITZ Energy and Environment, Raaba/Graz

Books and Reviews

- Gleis M et al. (2011) German experiences with thermal treatment processes. In: Seventh international congress & exhibition for energy efficiency & renewable for south east Europe, Sofia, 13–15 April 2011
- Grabda M et al. (2011) Waste to energy – practices and technologies in Japan. In: Seventh international congress & exhibition for energy efficiency & renewable for south east Europe, Sofia, 13–15 April 2011
- Thomé-Kozmiensky K (1993) Reaktoren zur thermischen Abfallbehandlung, EF – Verlag für Energie und Umwelttechnik GmbH, Berlin
- Thomé-Kozmiensky K (1994) Thermische Abfallbehandlung, EF – Verlag für Energie und Umwelttechnik GmbH, Berlin
- Wilén C et al. (2004) Finnish expert report on best available techniques in energy production from solid recovered fuels, Finnish Environment Institute, Helsinki

Wastewater Reclamation

IN S. KIM¹, BYUNG SOO OH¹, SEOKMIN YOON²,
HOKYONG SHON³, SANGHO LEE⁴, SEUNGKWAN HONG⁵

¹Center for Seawater Desalination Plant, Gwangju
Institute of Science and Technology (GIST), Buk-gu,
Gwangju, Republic of Korea

²Environment Research Division, Research Institute of
Industrial Science and Technology (RIST), Pohang,
Republic of Korea

³School of Civil and Environmental Engineering,
University of Technology, Sydney (UTS), Broadway,
NSW, Australia

⁴Department of Water Resources and Environmental
Research, Korea Institute of Construction Technology,
Goyang-Si, Gyeonggi-Do, Republic of Korea

⁵Department of Civil, Environmental & Architectural
Engineering, Korea University, Sungbuk-gu, Seoul,
Republic of Korea

Article Outline

Glossary
Definition of the Subject
Introduction
Describing Wastewater Reclamation
Technologies for Wastewater Reclamation
Future Directions
Bibliography

Glossary

Agricultural water Water used for crop production and livestock uses.

Concentrate (also called retentate) The portion of the feed stream that contains salts rejected from the membrane system. The term can also refer to the rejected mixed liquor suspended solids (MLSS) in a membrane bioreactor (MBR) process, which is recycled or wasted to maintain a given biomass concentration in the biological treatment process.

Conventional treatment Technologies such as activated sludge and trickling filters that remove BOD and suspended solids from wastewater. Conventional treatment technologies are often coupled with anoxic and anaerobic processes for nutrient removal.

Direct potable reuse Introduction of reclaimed water directly into a drinking water distribution system, without intervening storage or additional treatment.

Indirect potable reuse Augmentation of a raw water supply with reclaimed water followed by an environmental buffer. The mixture of raw and reclaimed water typically receives additional treatment before distribution as drinking water.

Membrane bioreactor (MBR) A process that combines a suspended growth activated sludge reactor with a membrane separation system. Membrane separation is accomplished by either microfiltration or ultrafiltration and used in place of conventional gravity sedimentation.

Pathogens Disease-causing organisms capable of inflicting damage on a host it infects.

Potable water Water deemed safe for human consumption, food preparation, and bathing.

Reclaimed water Municipal wastewater that has gone through various treatment processes to meet specific water quality criteria with the intent of being used in a beneficial manner.

Renewable water resources The water entering a country's surface and groundwater systems. Not all of this water can be used because some falls in a place or time that precludes tapping it even if all economically and technically feasible storage and diversion structures were built.

Secondary treatment Activated sludge treatment, commonly with nitrification, used for the removal of soluble organic matter and particulate constituents.

Trace organics Organic compounds detected at very low (minute) levels by the use of sophisticated instrumentation capable of measuring concentrations in the range of 10^{-12} to 10^{-3} mg/L.

Wastewater reclamation Treatment or processing of wastewater to make it reusable with definable treatment reliability and meeting appropriate water quality criteria.

Wastewater reuse The use of treated wastewater for a beneficial use, such as agricultural irrigation and industrial cooling.

Wastewater Used water discharged from homes, business, cities, industry, and agriculture. Various synonymous uses such as municipal wastewater (sewage), industrial wastewater, and stormwater.

Definition of the Subject

Wastewater can be defined as used and contaminated water, mainly through human activities, which is adversely affected in quality by anthropogenic influence. The meaning of reclamation is the process of artificially converting an unusable entity to one that is usable. Accordingly, the definition of wastewater reclamation is the process of reclaiming wastewater into a reusable form by artificial treatments. Through the treatment processes, the reclaimed water is able to meet appropriate water quality criteria so it can be returned to the environment to augment the natural systems from which it came, or reused for numerous purposes related to human activities. The reuse of reclaimed wastewater makes two important contributions. The first is to produce sustainable water resources, and the second is to reduce adverse environmental effects. Furthermore, if wastewater could be completely reclaimed, it would be possible to have zero discharges in the water recycle between humans and the natural environment.

Introduction

The total amount of water on the earth rarely changes due to external impacts, but the world now suffers from a scarcity of usable water, which is due to the dislocation of the earth's water cycle, with the main reason being climate change. It is true that rapid industrialization and population growth have been accelerating climate change and contributing to the increasing

scarcity of water around the world. Many countries are making great endeavor to overcome the water shortage problem by acquiring sustainable water resources via three main technologies; seawater desalination, rainwater harvesting, and wastewater reclamation. The properties of each of these technologies will be introduced in section "[Sustainable Water Resource](#)." All human activities use water, and this used water is discharged as wastewater (including sewage water and other discharged waters). The wastewater discharged from everywhere (e.g., homes, business, cities, industry, agriculture, and stormwater runoff) is treated in wastewater treatment plants to produce treated wastewater, called wastewater effluent, which will mix with natural water (e.g., river and lake). Target sources for wastewater reclamation include wastewater, wastewater effluent, and further treated wastewater effluent, but reuse of wastewater itself is significantly limited due to the water quality required for safe reuse. Technologies for wastewater reclamation are manifold according to the required purpose and water quality criteria. The applications of reclaimed water are categorized as follows: agricultural irrigation, landscape irrigation, industrial recycling and reuse, groundwater recharge, non-potable urban uses, and potable uses (see section "[Applications and Water Qualities for Using Reclaimed Water](#)"). The wastewater reclamation technologies can be divided into two types, conventional (processes installed in municipal wastewater treatment plants) and advanced (processes mainly related to membranes, such as membrane bioreactor, microfiltration, ultrafiltration, nanofiltration, and reverse osmosis for shifting wastewater to high-quality reclaimed water). For the most part, these methods are derived from physical, chemical, and biological treatment processes operated at municipal wastewater and drinking water treatment plants. The technical review of both conventional and advanced technology for wastewater reclamation will be discussed in sections "[Conventional Treatment Practice](#)" and "[Advanced Technologies for Producing High-Quality Reclaimed Water](#)." The keyword, "energy," is a critical issue around the world along with "water." Even though wastewater can be reclaimed as drinking water, it should be noted that the energy consumption of the applied technology is acceptable for its beneficial use. In the case of seawater desalination, a recent technical shift from thermal type to

reverse osmosis has been brisk because an energy-efficient plant is necessary for entering world markets. In section “[Energy Consumption: Comparison of Wastewater Reclamation with Desalination](#),” a comparative survey is presented for the energy consumption level between wastewater reclamation and seawater desalination. According to the applications for which the reclaimed water will be used, treatment goals are various.

Describing Wastewater Reclamation

Sustainable Water Resource

At global level, water source on Earth is abundant in nature, but the water source that is allowed to be accessed by humans is not enough. Over 97% of water on Earth is salt water; the remaining 5% is freshwater, including glaciers, groundwater, and surface water. Also, only 0.3% of the freshwater is surface water that can be easily used by humans. The global water cycle describes the continuous movement of water on Earth. Rivers flow into the oceans and seas, and the sun heats this water, which is the driving force of the water cycle. The heated water moves from oceans to land through a process of evaporation → condensation → precipitation, by which water becomes sustainable on land, supplying humans with sufficient water. However, the water cycle has changed due to global warming. Climate change has caused more frequent, severe rainfall and snowfall in some regions, and intensive drought in other regions, resulting in a significant reduction in the availability of renewable water resources for humans. In addition, the centralization of population into cities and industrial areas has made the use of various water sources difficult. About one third of the world's population (= about 6.8 billion in 2010) currently faces moderate to severe water shortages. An estimated 25% of the world population lacks access to clean drinking water and protection from waterborne diseases (US Census Bureau – World POPClock Projection). Recently, great efforts have been made to solve this water scarcity problem by finding alternative and sustainable water resources, which are classified into three main categories; rainwater harvesting, seawater desalination, and wastewater reclamation. First, a water storage device, such as a dam or reservoir, needs to be constructed that

can store water when it rains. Rainwater harvesting used to be the best way to conserve water from flowing into seas and oceans, but there are many constraints on this method. First of all, building a dam causes deterioration of the water quality. Furthermore, it is unsuitable for securing sustainable water because precipitation and the time for rain cannot be controlled. Second, desalination could be a solution for changing unusable (seawater) into available water (freshwater). In this case, seawater can be classified as an unlimited water resource. However, there are space restrains because this water resource only exists on the sea shore. Third, wastewater reclamation is the most popular way to produce usable water from wasted water. Wastewater reclamation could surpass the limitations of the other reclamation methods, because wastewater is produced everywhere human activities take place.

Applications and Water Qualities for Using Reclaimed Water

A summary of categories for the application of reclaimed water is given in [Table 1](#). The largest use of reclaimed water is for agriculture irrigation in most areas of the world. In 2002, about 46% of the total reclaimed water produced in California was used for agricultural irrigation (State of California 2003, 2004). Australia utilized 82% of their reclaimed water supply for agricultural irrigation in 2000 (Australian Bureau of Statistics 2004). In Israel, one of the most arid countries, it is estimated that 230 wastewater reclamation systems are in operation, recycling 72% of the municipal wastewater produced, primarily agricultural irrigation. The water supply from reclaimed water comprises about 15% of Israel's water recourses. In general, landscape irrigation is a major portion of the urban use of reclaimed water. For example, the Irvine Ranch Water District in southern California estimates that more than 70% of their total water use is for landscape irrigation. According to the data, the States of California and Florida are two largest users of reclaimed water in the United States, with golf course irrigation being major use of reclaimed water, comprising 50% and 36% of the total landscape irrigation use in each States, respectively. This indicates that landscape irrigation is a viable option for reducing the

Wastewater Reclamation. Table 1 Applications of reclaimed water [14]

Main category	Subcategory	Examples
Agricultural irrigation	Food crops	Crop grown for human consumption and consumed uncooked
	Nonfood crops and others	Fodder, fiber, seed crops, pastures, commercial plant nurseries, sod farms, and commercial aquaculture
Urban uses	Non-potable use	Landscape irrigation (parks, playgrounds, school yards, golf courses, cemeteries, greenbelts, and residential), fire protection, construction, ornamental fountains, road care/maintenance, car washing and in-building uses (toilets and air-conditioning)
	Potable use	Direct pipe to pipe water supply, blending with municipal water supply (reservoirs and groundwater)
Industrial uses	–	Cooling water, boiler feed, process water, construction activities, and washdown water
Environmental uses	Recreational use	Lakes and ponds used for swimming and snowmaking, fishing, and boating
	Environmental enhancement	Artificial wetlands, enhanced natural wetlands, sustained stream flows, and marsh enhancement
	Groundwater recharge	Groundwater replenishment, salt water intrusion control, and subsidence control

demand on potable water and decrease wastewater discharges to the environment. Industries commonly tend to use water at a relatively constant rate throughout the year compared to the seasonal and geographic variations associated with water use for agricultural and landscape irrigation. Therefore, a major benefit of using reclaimed water for industrial applications is that the water use requirements are consistent year-round. Other important considerations are the economic feasibility regarding system components and the related cost elements, such as reclaimed water supply, distribution system, onsite re-piping, engineering analysis of water quality and supplemental treatment alternatives, pretreatment capital and operating and maintenance costs, internal treatment needs, management of process residuals, including sludges and brines, as well as institutional, legal, and administrative activities. The contribution of reclaimed water for environmental enhancement has recently increased to satisfy the requirements of maintaining minimum water flows in rivers, streams, and wetlands, which is a major challenge in water resources management. The augmentation of various water resources using reclaimed water also increases the recreational opportunities.

When applying reclaimed water to its desired purposes, the water quality takes priority over

everything else. A review of the required water quality for the application of reclaimed water is discussed below.

Concerns on pathogen when using reclaimed water: Encouraging wastewater reclamation and ensuring public health from wastewater reuse have been serious issues since the 1930s. Improperly treated wastewater may increase the risk to public health by exposing humans to microbial pathogens, chemicals, and heavy metals. Therefore, the safe use of wastewater has long been discussed with respect to the quality of water for various types of reuse; urban, agricultural, industrial, recreational and environmental reuse, groundwater recharge, and indirect potable reuse. Of the risks associated with water reuse, microbial pathogens have been considered the most significant, because sewage from urban and agricultural areas are major sources, which contain many pathogens, including viruses, bacteria, protozoan parasites, and helminthes. G. Kamizoulis [10] investigated the number of pathogens in raw sewage, as summarized in Table 2. Accordingly, thermotolerant coliforms (fecal coliforms) and enteric viruses were found to be the most abundant in raw sewage. It is clear that these pathogens cause various diseases to humans when they are exposed, either directly or indirectly, to the untreated sewage.

Wastewater Reclamation. Table 2 Typical numbers of excreted organisms in raw sewage [10]

Organisms	Numbers in wastewater (per L)
Bacteria	
Thermotolerant coliforms	10^8 – 10^{10}
<i>Campylobacter jejuni</i>	10 – 10^4
<i>Salmonella</i> spp.	1 – 10^5
<i>Shigella</i> spp.	10 – 10^4
<i>Vibrio cholerae</i>	10^2 – 10^5
Intestinal helminths	
<i>Ascaris lumbricoides</i>	1 – 10^3
<i>Ancylostoma/Necator</i> 03	1 – 10^3
<i>Trichuris trichiura</i>	1 – 10^2
Protozoa	
<i>Cryptosporidium parvum</i>	1 – 10^4
<i>Entamoeba histolytica</i>	1 – 10^2
<i>Giardia intestinalis</i>	10^2 – 10^5
Viruses	
Enteric viruses	10^5 – 10^6
Rotavirus	10^2 – 10^5

Although many microbial pathogens exist in the raw sewage, the survival times outside of their hosts are limited, even without any disinfection process. Bradley (1983) investigated the survival time of pathogens at 20–30°C when exposed to environmental media, including freshwater and sewage, crops and soil (Table 3). According to the results, the survival time was shortest when the pathogens are in crops. Meanwhile, helminthes have been shown to survive for several months in water and soil.

Traditionally, coliforms and fecal coliforms are commonly used as pathogen indicators due to their large numbers in the feces of humans and warm-blooded animals, easy and rapid detection, and lack of growth in natural habitats outside of their hosts [7]. For the safe use of wastewater, the first guideline was enacted for irrigation in California, USA, in 1918 (California Table 22), which adopted a strict guideline for coliforms and fecal coliforms because California had advanced water treatment technology, but little

Wastewater Reclamation. Table 3 Survival times of pathogen at 20–30°C [6] (unit: days)

Pathogen	Freshwater and sewage	Crops	Soil
Viruses	<120 (<50)	<60 (<15)	<100 (<20)
Bacteria	<60 (<30)	<30 (<15)	<70 (<20)
Protozoa	<30 (<15)	<10 (<2)	<70 (<20)
Helminth	Many months	<60 (<30)	Many months

Figures in brackets show the normal survival time

available epidemiological evidences. The USEPA later developed federal guidelines on the basis of the treatment process, and finally published “Guidelines for water reuse” in 1992. The USEPA guidelines contain reclaimed water quality guidelines, with recommended monitoring and setback distances. The USA developed a water treatment process to meet the strict water quality, such as coagulation/flocculation, filtration, and disinfection, and these standards were also used for the safe quality of imported crops. In 1973, WHO announced the first wastewater reuse guidelines, with an epidemiological approach, which contained values for tolerable coliforms and fecal coliforms for the practical use of reclaimed wastewater in developing countries where sewage has been historically used for irrigation of crop without an appropriate treatment. The first guidelines have been revised with more epidemiological evidences, treatment processes, and irrigation systems, and WHO finally published the current guidelines in 1989, titled “Health Guidelines for the Use of Wastewater in Agriculture and Aquaculture.” For the development of standards in the USA, California Table 22 influenced the regulation program of each state, but each state has different standards with different approaches according to the treatment process or water quality. A range of standards is now in use, as shown in Table 4. The USA standard consists of treatment technology, BOD5, TSS, Turbidity, and coliforms (total and fecal), with ten types of water use; urban reuse (restricted and unrestricted), agricultural (food crops and nonfood crops), recreational (restricted and unrestricted), environmental (wetland), industrial, groundwater recharge (non-potable aquifer), and

Wastewater Reclamation. Table 4 Examples of microbial quality standards used by various States in the USA [3]

Exposure route	Total coliforms per 100 mL		Fecal coliforms per 100 mL		Enteric viruses per 40 L	
	<i>n</i> ^a	Range of values	<i>N</i>	Range of values	<i>N</i>	Range of values
Spray irrigation ^b	4	2.0–100	3	2.2–200	1	1
Surface irrigation ^b	2	100	9	10–1,000	0	–
Parks and playgrounds ^c	8	2.2–100	3	10–100	1	128
Golf courses and open space ^d	6	2.2–1,000	5	0–100	0	–

^aNumber of states involved out of the 13 selected^bIncludes food crop irrigation^cIncludes playgrounds^dIncludes cemeteries^eArizona is the only state that has a virus standard**Wastewater Reclamation. Table 5** Recommended microbiological quality guidelines for wastewater use in agriculture^a

Reuse conditions	Exposed group	Intestinal nematodes (/L ^{a,e})	Fecal coliforms (/100 mL ^{b,e})	Wastewater treatment expected to achieve required quality
Irrigation of crops likely to be eaten uncooked, sports fields, public parks ^c	Workers, consumers, public	<1	<1,000	A series of stabilization ponds designed to achieve the microbiological quality indicated, or equivalent treatment
Irrigation of cereal crops, industrial crops, fodder crops, pasture, and trees ^d	Workers	<1	None set	Retention in stabilization ponds for 8–10 days or equivalent helminth removal
Localized irrigation of crops if category B exposure of workers and the public does not occur ^e	None	n/a	n/a	Pretreatment as required by the irrigation technology, but not less than primary sedimentation

^aArithmetic mean^bGeometric mean^cIn specific cases, local epidemiological, sociocultural, and environmental factors should be taken into account, and the guidelines modified accordingly^d*Ascaris* and *Trichuris* species and hookworms^eDuring the irrigation period^fA more stringent guideline (200 fecal coliforms/100 mL) is appropriate for public lawns with which the public may come into direct contact^gIn the case of fruit trees, irrigation should cease 2 weeks before the fruit is picked and none should be picked off the ground

indirect potable reuse. The USA guidelines and standards influenced those of Australia, South Korea, and so on (Mehlika 1996; Ursula 2000; [9]), and the WHO guidelines affected European and Mediterranean nations in establishing their regulations. Particularly, a European level guideline has not been established, but is now being developed through the AQUAREC project [1].

Water quality of wastewater reuse for agriculture: Agricultural irrigation accounts for the largest part of

water demand, approximately 80% of the worldwide; therefore, it also will also become the major water reuse application [21]. As describe before, the first standards are for microbiological parameters that show the presence of pathogens in wastewater. Water quality for irrigation in agriculture has been subdivided into three groups by the WHO (1989), as shown in Table 5; Group A (Irrigation of crops to be eaten uncooked, sports fields, and public parks), Group B (Irrigation of

cereal crops, industrial crops, fodder crops, pasture, and trees) Group C (exposure of workers and the public does not occur), in which the exposed group, helminthes, and fecal coliforms are described separately. The fecal coliforms in reclaimed water for uncooked crops are below 1,000 cfu/100 mL. The USEPA [20] guidelines for treated wastewater in agriculture are shown in Table 6, with type of use subdivided into three groups; similar to WHO (1989); Food crops not commercially processed, Food crops commercially processed, for nonfood crops. Contrary to the WHO guideline, zero fecal coliforms in 100 mL are required for irrigation water quality of uncooked crops, in addition to BOD₅, pH, SS, and residual chlorine. It is also recommended for reclaimed water monitoring and setback distances for each type of use. In the USA, each state has different microbiological standard, with different approach according to the treatment process or water quality, and a range of standards is in use. For unrestricted irrigation of food crops, the range is 2.2–20 fecal coliforms bacteria/100 mL, and for restricted irrigation is 23–200 fecal coliforms bacteria/100 mL.

Water quality of wastewater reuse for other type of applications: The USEPA standard for water quality (1992) describes the standards for each state for ten types of water use. As with water quality for agriculture, the regulatory items include treatment method, BOD₅, TSS, Turbidity, and coliforms (total and/or fecal coliforms). In 1998, WHO published “Guidelines for Safe Recreational Water Environments.” These guidelines deal with fecal coliforms’ contamination of recreational water, with dose–response results used to develop criteria linking the microbial concentration in recreational waters to gastroenteritis. Table 7 shows the water quality for reclaimed wastewater and the qualitative requirements of eight types of water use.

Other water quality parameters: Besides microbial pathogens, other water quality parameters related with wastewater reuse are toxic metal accumulation and the salinity of wastewater. The availability and accumulation of heavy metals in crops depends on soil factors, such as pH, cation exchange capacity (CEC) and moisture content, and plant factors, such as species variety, plant parts used for consumption, and temperature. The salinity of wastewater is a problem in some Mediterranean countries, as it exerts an osmotic effect on

plant growth. The USEPA updated the draft regulations in July 2003, requiring monitoring of emerging pollutants of concerns (EPOCs). The draft regulations require annual monitoring for pharmaceuticals, endocrine disrupting chemicals, and other chemical indicators in municipal wastewater [21].

Technologies for Wastewater Reclamation

Conventional Treatment Practice

Conventional wastewater treatment, which is also applied for water reuse, consists of a combination of physical, chemical, and biological processes, with operations to remove solids, organic matter and, sometimes, nutrients. Different degrees of treatment are used to achieve the desired levels of contaminant removal for water reuse, including preliminary, primary, secondary, tertiary, and disinfection, as well as advanced treatment.

1. *Preliminary treatment:* Preliminary treatment removes large solid materials from the raw wastewater before they are able to damage or clog the pumps and skimmers of the primary treatment clarifiers. Screening and grit removal are typical examples of preliminary treatment.
2. *Primary treatment:* Primary treatment aims to remove settleable organic and inorganic materials by sedimentation and/or skimming. It is effective at removing large fractions of the total suspended solids, oils, and greases. Some organic nitrogen, organic phosphorous, and heavy metals associated with those solids are also removed, but colloidal and dissolved constituents are not removed. For nonfood crop irrigation, primary treatment may be a sufficient treatment. However, secondary treatment may be required to prevent potential adverse effects. Sedimentation tanks or clarifiers are generally used as a primary treatment. The sludge is processed by methods such as aerobic digestion, storage in sludge lagoons, land application, and others.
3. *Secondary treatment:* Secondary treatment is additional to the primary treatment of effluent to remove the residual organics (as indicated by BOD) and suspended solids. In most cases, secondary treatment involves aerobic biological treatment

Wastewater Reclamation. Table 6 US-EPA/USAID guidelines for agricultural reuse of wastewater [20]

Type of reuse		Treatment	Reclaimed water quality	Reclaimed water monitoring	Setback distance
Urban reuse					
Landscape irrigation, vehicle washing, toilet flushing, fire protection, commercial air conditioners, and other uses with similar access or exposure to the water		Secondary ^a	pH = 6–9	pH: weekly	50 ft (15 m) to potable water supply wells
		Filtration ^b	BOD < 10 mg/L	BOD: weekly	
		Disinfection ^c	Turbidity < 2NTU ^d	Turbidity: continuous	
			FC = zero coliform/100 mL ^e	Coliform: daily	
			Cl ₂ residual = 1 mg/L min	Cl ₂ residual: Continuous	
Agricultural reuse	Food crops not commercially processed				
	Surface or spray irrigation of any food crop, including crops eaten raw	Secondary	pH = 6–9	pH: weekly	50 ft from potable water supply wells
		Filtration	BOD ≤ 30 mg/L	BOD: weekly	
		Disinfection	Turbidity < 1NTU ^d	Turbidity: daily	
			FC = zero /100 mL ^e		
			Cl ₂ residual = 1 mg/L min	Coliforms: daily	
				Cl ₂ residual: continuous	
	Food crops not commercially processed				
	Surface irrigation of Orchards and Vineyards	Secondary	pH = 6–9	pH: weekly	100 ft from areas accessible to public
		Disinfection	BOD ≤ 30 mg/L	BOD: weekly	
			SS = 30 mg/L	SS: daily	
			FC ≤ 200/100 mL	FC: daily	
			Cl ₂ residual = 1 mg/l min	Cl ₂ residual: continuous	
	For nonfood crops				
	Pasture for milking animals; fodder, fiber and seed crops	Secondary	pH = 6–9	pH: weekly	300 ft (90 m) to potable water supply wells
		Disinfection	BOD ≤ 30 mg/L	BOD: weekly	
			SS < 30 mg/L	TSS: daily	
			FC < 200/100 mL ^d	Coliforms: daily	
			1 mg/L Cl ₂ residual (min)	Cl ₂ residual: continuous	

Wastewater Reclamation. Table 6 (Continued)

Type of reuse	Treatment	Reclaimed water quality	Reclaimed water monitoring	Setback distance
<i>Indirect potable reuse</i>				
Groundwater recharge by spreading into potable aquifers	Site specific	Site specific	pH: daily	100 ft (30 m) to areas accessible to the public (if spray irrigation) site specific
	Secondary and disinfection (min)	Meet drinking water standards after percolation through vadose zone	Turbidity: continuous	
	May also need filtration and/or advanced wastewater treatment		Coliforms: daily	
			Cl ₂ residual: continuous	
			Drinking water standards: quarterly	
			Other: depends on constituent	

Source: U.S. Environmental Protection Agency [20].

^aSecondary treatment processes include activated sludge processes, trickling filters, rotating biological contactors, and many stabilization pond systems. Secondary treatment should produce effluent in which both the BOD and TSS do not exceed 30 mg/L

^bFiltration means passing the effluent through natural undisturbed soil or filter media such as sand and/or anthracite

^cDisinfection means the destruction, inactivation, or removal of pathogenic microorganisms. It may be accomplished by chlorination, or other chemical disinfectants, UV radiation, or other processes

^dThe number of fecal coliforms organisms should not exceed 800/100 mL in any sample

^eThe number of fecal coliforms organisms should not exceed 14/100 mL in any sample

processes. High-rate biological processes, such as activated sludge processes, trickling filters or biofilters, oxidation ditches, and lagoons, as well as rotating biological contactors (RBC), are commonly used. A combination of two of these processes in series can be used to treat municipal wastewater containing a high concentration of organic material from industrial sources.

High-rate biological processes are characterized by relatively small reactor volumes and high concentrations of microorganisms compared with low-rate processes. The microorganisms must be separated from the treated wastewater by sedimentation to produce clarified secondary effluent. High-rate biological treatment processes are

effective at removing BOD and suspended solids, as well as some of the heavy metals. However, they remove very little of the phosphorous, nitrogen, nonbiodegradable organics, or dissolved minerals.

4. *Tertiary treatment:* Tertiary treatment is employed when specific undesirable wastewater constituents cannot be removed by secondary treatment. These constituents include nitrogen, phosphorous, additional suspended solids, refractory organics, and heavy metals and dissolved solids.
5. *Disinfection:* Disinfection is intended to reduce the number of pathogenic microorganisms in the water. Chlorine is normally used, but ozone and ultraviolet irradiation can also be used to meet advanced wastewater treatment requirements.

Wastewater Reclamation. Table 7 Consumptive user of reclaimed wastewater and qualitative water quality requirements [13]

Consumptive use	Removal of pathogens	Chlorine residual or other disinfection	Removal of solids & turbidity	Presence of dissolved oxygen	Removal of BOD & COD	Removal of nutrients	Removal of taste, odor and color	Removal of trace organics & metals	Removal of excess salinity
Landscape	*	—	*	*	*	—	*	*	0
Nonfood crops	**	*	*	*	**	—	*	*	0
Food crops	***	***	***	**	***	—	**	**	*
Groundwater recharge	***	**	***	**	***	***	***	***	—
Industrial uses	**	**	***	***	***	***	***	**	**
Urban non-potable uses	****	****	****	****	****	***	****	****	**
Potable uses	*****	*****	*****	****	*****	****	*****	*****	***

(—) no need, (0) usually not essential, (*) slight need, (**) moderate need, (***) strong need, (****) stringent requirements, (***** very stringent

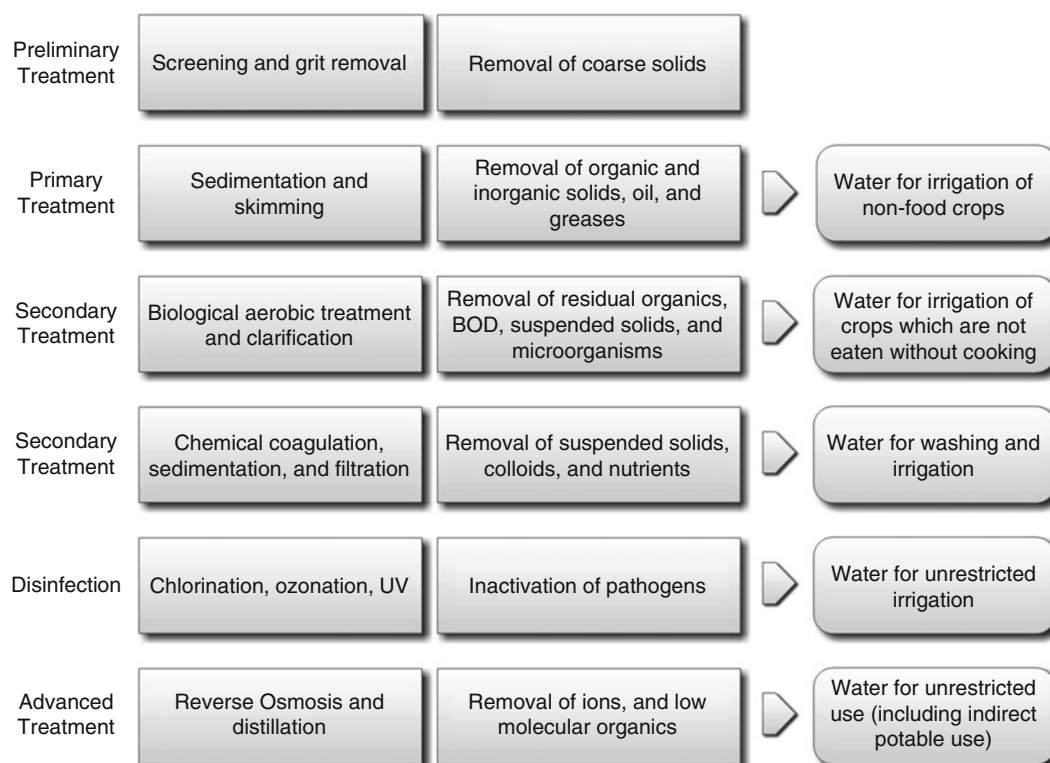
6. *Advanced treatment:* Advanced treatment includes membrane technology (including ultra/nanofiltration or reverse osmosis) as well as water remineralization. The advanced technologies have gained interest for unrestricted uses of water because they are now much cheaper than in the past. This will be described in the next section (Fig. 1).
7. *Sludge treatment and disposal:* The sludge accumulated in a water reuse process must be treated and disposed of in a safe and effective manner. The most common treatment options are anaerobic digestion, aerobic digestion, and composting. Incineration is less common due to air emissions concerns. Anaerobic digestion is widely used for the treatment of domestic sewage. One major feature of anaerobic digestion is the production of biogas (with the most useful component being methane), which can be used for electricity production and/or heating purposes.

Advanced Technologies for Producing High-Quality Reclaimed Water

The increased need to secure a high-quality water supply from reclaimed water has resulted in the emergence of advanced technologies, including membrane bioreactors (MBR), nanofiltration (NF), reverse osmosis (RO), ion exchange (IEX), distillation, chemical oxidation, and carbon adsorption.

1. *Membrane bioreactor (MBR):* Membrane bioreactors combine activated sludge treatment with a membrane separation process. The reactor is operated in a similar way to conventional activated sludge, but a low pressure membrane (microfiltration or ultrafiltration) is used to perform the sludge separation. This allows the production of high-quality water in only a one unit operation.

The MBR process has been shown to provide high-quality effluent, with high BOD removal and



Wastewater Reclamation. Figure 1
Treatment technologies for water reuse

complete TSS reduction. Depending on the design, configuration, and need for nutrient removal, the activated sludge portion of the process can provide significant denitrification and phosphorus reduction. The MBR effluent has the added advantage of having low turbidity, making it possible for use as feed water to reverse osmosis (RO) in industrial or indirect potable reuse systems. Moreover, the overall footprint of an MBR system is much smaller than a conventional activated sludge facility.

Membrane configurations can be classified as: (1) sidestream and (2) submerged MBR. A sidestream MBR pumps sludge from an activated sludge reactor to a pressure-driven membrane for solid-liquid separation. A submerged MBR has membrane modules submerged in the bioreactor and operates under vacuum pressure. The submerged MBR configurations are preferred to the sidestream ones, especially for reuse of domestic wastewater. The energy demand of the submerged MBR can be up to two orders of magnitude lower than that of the sidestream systems.

MBR systems can be sited close to the points of potential reuse. Thus, local needs for reclaimed water are met more economically because of the reduced requirements for pipelines, pumping, and storage. Moreover, the product water quality is suitable for a variety of reclaimed water applications and, after disinfection, can be used for a number of non-potable, unrestricted uses.

During the operation of an MBR, the rejected constituents in the retentate tend to accumulate at the membrane surface, producing various phenomena, which lead to a reduction in the flow of water through the membrane (i.e., the flux) at a given transmembrane pressure (TMP) or, conversely, an increase in the TMP for a given flux (reducing the permeability, which is the ratio of flux to TMP). These phenomena are collectively referred to as fouling. Since membrane fouling represents the main limitation to membrane process operation, MBR systems are designed to mitigate fouling problems.

MBR systems using low pressure membranes cannot retain organic species and ions. To overcome these limitations, high retention membrane bioreactors (HRMBRs) have recently been

investigated. HRMBRs use high rejection membranes, such as nanofiltration (NF), forward osmosis (FO), and membrane distillation (MD), in place of MF/UF. This allows a combination of the conventional MBR and high rejection membrane separation in a single step. However, this technology is still in its early stage of development, and many challenges, such as low recovery and accumulation of ions in the bioreactors, still have to be resolved.

2. *Nanofiltration/reverse osmosis*: With the increased use of reclaimed water for applications where quality and reliability are critical, such as indirect potable reuse and some industrial uses, the increased removal of dissolved solids and trace constituents may be required. In addition, dissolved solids may cause scaling or corrosion of equipment and piping systems, especially in cooling tower systems. Membrane technologies, such as nanofiltration (NF) and reverse osmosis (RO), can be applied to remove these constituents.

NF and RO are capable of separating dissolved ions from the feed stream, and require hydrostatic pressure to overcome the osmotic pressure of the feed stream. The key difference between NF and RO is the degree of removal of monovalent ions, such as sodium and chloride ions. Reverse osmosis removes monovalent ions in the 98–99% range, while removals with NF membranes vary between 50% and 90%, depending on the material and manufacture of the membrane. With the development of new low-pressure membranes, RO is finding increased use in water reuse applications.

NF is often used for removing salts to meet TDS requirements for groundwater recharge and for water softening applications. NF is also used in conjunction with RO to remove divalent and trivalent ions, which enable a higher recovery rate with RO. However, NF may not replace RO in reuse applications of municipal wastewater because of its low rejection of monovalent ions.

RO is effective at reducing TDS for unrestricted water reuse, including indirect portable use and power plant cooling water, and is also used to remove trace organics, such as endocrine disrupting chemicals and pharmaceuticals. However, RO is relatively ineffective at retaining non-charged small organic molecules (i.e., urea). Moreover, RO

requires high pressure to produce water, leading to a high energy cost.

Like all membrane processes, NF and RO membranes can be fouled by various constituents in the feed stream. Thus, pretreatment prior to the NF/RO process requires the removal of suspended particulates, such as colloidal material. This type of removal will normally require the use of a coagulant/filtration process, acid addition, and/or filtration, such as microfiltration (MF) or ultrafiltration (UF).

Another drawback for the implementation of NF/RO processes is the need for additional treatment of the concentrate. Membrane concentrate is primarily concentrated feed raw water materials, plus any added chemicals used for pretreatment purposes. Because of the high levels of dissolved solids, together with other toxic compounds in the raw water, the discharge of untreated membrane concentrate poses a significant risk to the environment. Various options exist for the disposal of concentrate from NF/RO plants, including discharge to surface water, discharge to wastewater treatment plants, deep wells, land application, evaporation ponds, and wastewater evaporators. Management of NF/RO residuals requires evaluation of the available management alternatives, regulations, costs and site-specific conditions.

3. *Electrodialysis*: Electrodialysis (ED) is an electrically driven process that separates mineral salts and other species using direct electrical potential as the driving force. When a direct current is passed through a solution, the positive and negative ions migrate to the negative and positive electrodes, or the cathode and anode, respectively. Accordingly, ED is not as effective at removing microorganisms and many anthropogenic organic compounds. Electrodialysis and electrodialysis reversal (a modification of the electrodialysis process) are used principally for desalting brackish water, and have also been used for demineralizing reclaimed water.
4. *Ion exchange*: Ion exchange is a process used for the removal of dissolved ionic constituents, where ions of a given species are displaced from a solid phase material by ions of a different species from the solution. This is applied for demineralization of

water, removal of specific constituents (i.e., nitrate), and softening.

Residual organic matter found in biological treatment effluents can cause blinding of the ion exchange beds, resulting in high pressure loss and inefficient operation, for example, where ion exchange is used following filtration. Therefore, some form of chemical treatment and clarification is required before ion exchange treatment. Moreover, ion exchange resins should be regenerated after a certain period of time, resulting in the challenge of brine management. The disposal of high TDS regeneration brine remains the primary obstacle to the wider use of ion exchange technology.

5. *Distillation*: Distillation is a unit operation, where the components of a liquid solution are separated by vaporization and condensation, which involves multiple effect distillation (MED), multistage flash evaporation (MSF), and vapor compression distillation (VCD). The principal issues with the application of distillation processes for water reclamation are the high energy input required to evaporate the feedwater, carryover of volatile constituents found in treated reclaimed water, and the degree of subsequent cooling and treatment that may be required to renovate the distilled water. In addition, all distillation processes reject part of the influent feedwater and produce a concentrated waste stream that must be managed. Scaling and corrosion are also problematic.
6. *Chemical oxidation*: Chemical oxidation is applied for the destruction of anthropogenic or synthetic toxic organics in water reuse systems. Oxidants used in water reclamation are chlorine, ozone, chlorine dioxide, permanganate, hydrogen peroxide, and hydroxyl radicals. The principal applications of conventional chemical oxidation in water reclamation are for odor control, hydrogen sulfide control, color removal, iron and manganese removal, disinfection, control of biofilm growth, and biofouling in treatment process and distribution system components, as well as the oxidation of selected trace organic constituents.

The major concern with any chemical oxidation process is the potential for the formation of toxic by-products due to incomplete oxidation. Oxidation processes using ozone can produce brominated

byproducts, as well as bromates, in water containing bromide ions. High concentrations of carbonate and bicarbonate in some reclaimed water can react with hydroxyl radicals and reduce the efficiency of oxidation processes. Metals ions and residual organics may also affect the performance of oxidation processes.

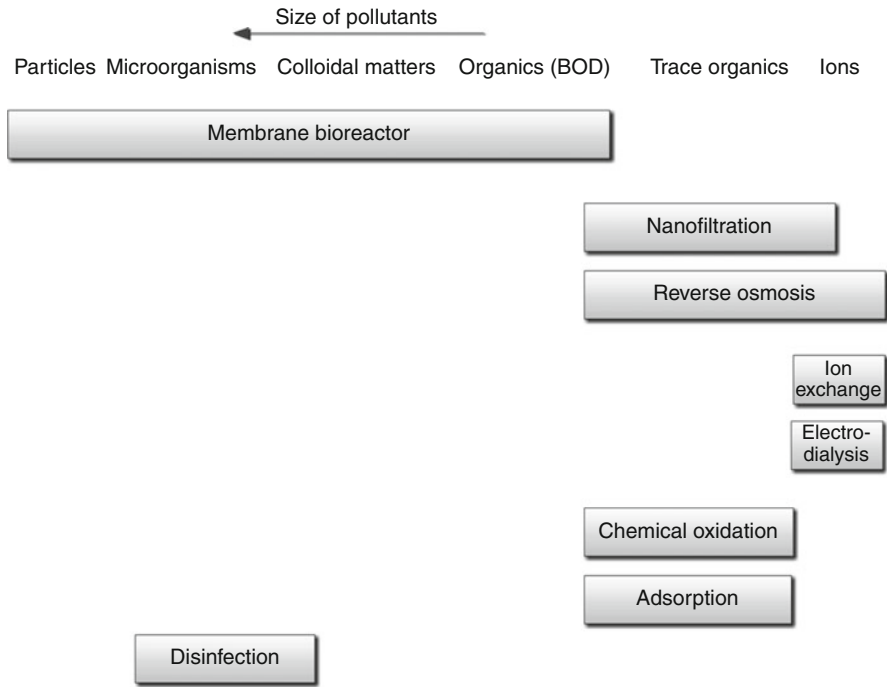
- 7. *Adsorption:* In water reclamation, adsorption is used for the removal of substances in solution by their accumulation on a solid phase. Adsorption treatment of reclaimed water is usually thought of as a polishing process for water that has already received normal biological treatment. Adsorption has been used for the removal of refractory organic constituents, residual inorganic constituents, such as nitrogen, sulfides and heavy metals, and for odorous compounds.

The adsorption process in water reuse applications is limited by several factors; the relatively high capital cost due to the area requirements for the carbon contactors, and waste adsorbent due to the difficulty in its regeneration, which may require disposal as a hazardous waste due to the presence of toxic constituents (Fig. 2).

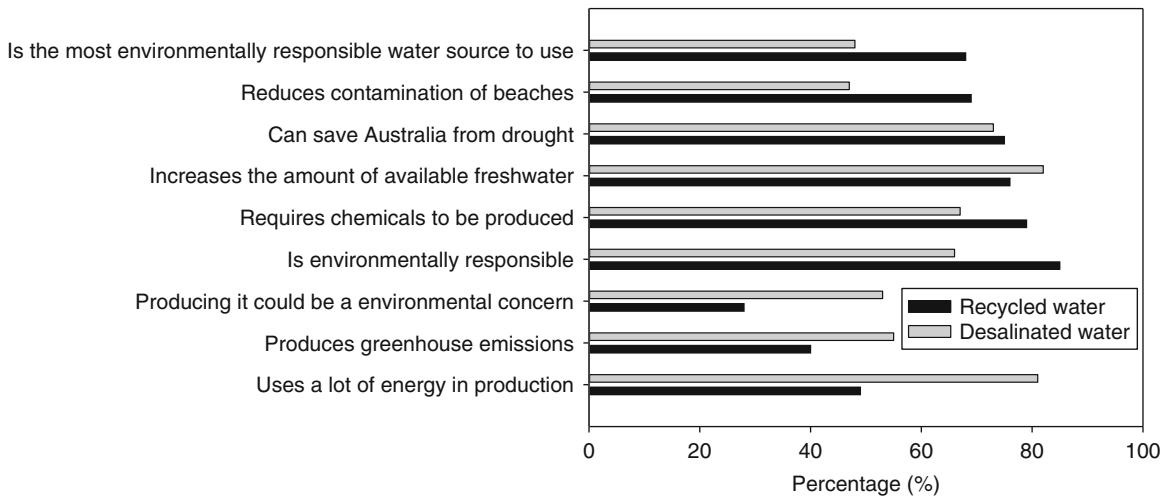
Energy Consumption: Comparison of Wastewater Reclamation with Desalination

Due to the high costs of power and an increasing concern over greenhouse gas emissions, the issue of energy reduction in wastewater treatment and reuse is becoming increasingly important. For instance, in California USA, an energy-efficiency program with the California Energy Commission is helping the state to achieve its goals of reducing per person water use by 20%, increasing renewable energy generation by 33%, and decreasing greenhouse gas emissions by 30% by 2020 [8], which could reduce the energy consumption in water/wastewater treatment plants by up to 30%.

The cost factors associated with wastewater treatment are pretreatment, chemical addition, cleaning, maintenance, and capital. In addition, water treatment and delivery require energy, which is another essential parameter in deciding on a wastewater treatment technology because it is a large fraction of the cost of water provision. Dreizin [5] reported that the energy cost was 25% and 65% of the total and operational costs, respectively, meaning that energy is the decision-making



Wastewater Reclamation. Figure 2
Advanced treatment for constituents in wastewater



Wastewater Reclamation. Figure 3

Comparative perceptions/knowledge about environmental issues of recycled and desalinated water in Australia [4]

parameter with regards of the economics of different source waters. In general, the specific energy consumption for surface, brackish, or wastewater is 0.4–1.0 kWh/m³; whereas, that for seawater is 3–3.4 kWh/m³. As treatment systems advanced, the energy requirements will become higher. Especially, membrane technology is recognized as a cost-intensive process in water/wastewater treatment. These membrane processes, i.e., RO, can be applied to wastewater reuse and seawater desalination, which are the most suitable alternative water resources; therefore, these two processes are often compared in terms of energy consumption. In this section, the energy consumption of membrane-based solutions for wastewater reuse and desalination are compared.

Public perception of energy consumption for recycled versus desalinated water: Figure 3 shows a survey of the comparative perception/knowledge about environmental issues of recycled and desalinated waters [4], which was carried out by 1,000 respondents to a 30 min online questionnaire conducted in Australia. Health, environmental, and cost concerns were found to be the major issues. As expected, wastewater reuse is primarily perceived as a health issue, while the concerns about desalinated water are the environmental and cost issues. Many people recognize that desalinated water uses more energy and produces greenhouse emissions.

Wastewater Reclamation. Table 8 Comparison of energy and O&M costs for recycled water [11, 18]

	Pretreatment		Pretreatment-RO	
	CAS	MF	CAS-RO	MF-RO
Utilities (kWh/m ³)				
Power	0.2	0.2	1.3	0.5
O&M costs (US cents/m ³)				
Power	2	1.8	13.2	5.3
Other	14	3.2	12.7	10.6
Total	13	5	25.9	15.9

Thus, it is essential to evaluate the detailed comparative energy costs between recycled and desalinated waters.

Energy cost of recycled water: The relative contributions of surface water, groundwater, desalinated water, and recycled water to the total water abstracted for all uses for humans are 63.2%, 36.3%, 0.29%, and 0.19%, respectively [12]. Recycled water has the lowest percentage, possibly due to the strict public acceptance of health concerns. However, the quality of recycled water is as good as that of desalinated water, as both technologies utilize RO-based membranes.

Wastewater Reclamation. Table 9 Energy consumption for recycled and desalinated water [15]

Water	Total dissolved solid (mg/L)	CAS (kWh/m ³)	Pretreatment (kWh/m ³)	RO system (kWh/m ³)	Total treatment (kWh/m ³)
Surface water	50–300				0.15–0.3
Wastewater	300–800	0.3–0.6	0.1–0.2	0.4–0.5	0.8–1.0 (1.3)
Wastewater MBR	300–800		0.8–1.0	0.4–0.5	1.2–1.5
Brackish water	900–2,200		0.1–0.3	0.6–0.9	0.8–1.0
Seawater	38,000		0.3–1.0	2.0–3.0	2.3–4.0

Wastewater Reclamation. Table 10 Cost comparison for producing water from CAS and seawater (Cote et al. 2005)

Component	Units	CAS	Seawater	Ratio (seawater/CAS)
Capital costs				
Infrastructure and pretreatment	US\$/m ³ /day	161	320	1.9
RO	US\$/m ³ /day	321	624	1.9
Total	US\$/m ³ /day	482	944	2.0
Total life cycle costs				
Capital	US\$/m ³	0.07	0.25	3.4
O&M	US\$/m ³	0.21	0.37	1.8
Total	US\$/m ³	0.28	0.62	2.2

Treatment of recycled water based on RO is divided into two processes: (1) conventional activated sludge (CAS) followed by MF/UF-RO, and (2) MBR-RO ([15]; Cote et al. 2005). The CAS in wastewater treatment is known as a biological process followed by clarification, but the CAS effluent is not appropriate as the feed water for RO due to severe membrane fouling. This leads to a combination of MF/UF and then on to RO. Prior to the MF/UF, lime clarification and sand filtration are used, which result in high chemical costs. Thus, the integrated MBR-RO systems for recycled water have been favorably used in many wastewater reuse plants. Integrated systems consist of screening, MBR, and RO, and do not require complex pretreatment steps, as required for CAS-MF/UF-RO. Reardon et al. [18] reported that the energy costs of the CAS versus MF pretreatment were similar, while the O&M costs of the CAS pretreatment were approximately three times higher than those of MF. On the other hand, when incorporated with RO, the energy

consumption of the CAS-RO system was 2.6 times higher than that of the MF-RO. This suggests that the MBR-RO has various advantages and significant potential for wastewater reuse (Table 8).

Energy comparison of recycled versus desalinated water: The major issue of recycled versus desalinated water is energy, which plays an important role in deciding on alternative water resources from the areas suffering from water shortage. Desalination has been regarded as a cost-intensive process due to the high energy requirement. Assuming the cost of electricity as \$0.10/kWh, for RO units using power generators, the cost of power would be \$0.30/kWh based on 3 kWh/L (assumed diesel cost of \$0.90/L). Compared to desalination, recycled water costs less due to: (1) no need for coagulant addition, (2) single pass RO, but 2-pass RO for desalination, (3) higher RO recovery, (4) higher RO flux, and (5) lower RO feed pressure (Cote et al. 2005).

Pearce [15] summarized the energy usages for various water and wastewater resources (Table 9).

Cote et al. (2005) reported a cost comparison for producing water from CAS and sweater (Table 10). Both surveys showed that desalinated water was 2–3 times more expensive than recycled water in terms of energy usage and water production costs. The key parameters of cost and energy usage differ between recycled and desalinated water due to the salinity of the latter. When the salinity increases, (1) the addition of chemical coagulant is needed, (2) smaller recovery (50%) from seawater (75% for wastewater), which leads to a larger pretreatment system, and (3) higher pressure due to high osmotic pressure with high salinity. This indicates that the difference (2–3 times) in the energy consumption and water production costs for recycled and desalinated water is unlikely to change as the difference in costs is generated due to the nature of saline water.

Future Directions

In recent years, the MBR process has become a viable alternative technology for wastewater treatment. Research on the MBR process has increased exponentially over the last decade, and is expected to keep increasing in the future. Despite rigorous and active research activities, membrane fouling still remains as a major obstacle for the practical and wide application of MBR technology. Furthermore, the great demand for wastewater reclamation and reuse due to the shortage of pristine water resources all over the world has pushed this technology to evolve into a more efficient process for controlling various micro-contaminants, in addition to the well-known biological nutrients, N and P. In this section, based on the current status of the knowledge obtained from the recent literature, the future directions of research and development on MBR technology are briefly discussed.

Various factors affecting fouling of the MBR process have been investigated and documented in numerous publications. However, many operational MBR plants have suffered from deterioration in the membrane productivity, mainly due to fouling. Thus, a fundamental understanding of membrane fouling mechanisms will continue to be the main focus of future research, with even more emphasis on enhancing the economic feasibility of this technology. More specifically, sophisticated analytical techniques will be developed to identify and characterize various chemical and

biological components of membrane foulants. For example, the interactions between foulants and membrane surfaces or pores could be directly measured using various tools, including AFM, to determine the fouling potential of these components. Current lab-scale analytical technologies need to be further developed, refined, and simplified to be able to monitor fouling phenomena in real applications. In situ monitoring techniques will become more significant for the effective operation of full-scale MBR plants in the future.

In addition to fundamental assessment and practical monitoring of membrane fouling, more scientifically designed fouling control methods are expected to be developed in the future, which will be made possible by the elucidation of fouling mechanisms. A novel method utilizing a biological mechanism control, known as Quorum Sensing, is one example. Biological and chemical methods developed on the basis of fouling mechanisms should be easier and cheaper to apply to large-scale plants. Lastly, new, improved membrane modules for the control of fouling during MBR processes will continue to develop and reduce the capital costs, and enhance the hydrodynamic conditions.

The MBR technology was initially developed based on MF/UF membranes capable of achieving solid-liquid separation for SS and coliforms. However, environmental concern regarding the protection of natural water resources put more pressure on the development of MBR processes that were more suitable and effective for the control of biological nutrients, such as nitrogen and phosphorus. High nitrification can be easily accomplished under typical MBR processes operating conditions, characterized by high MLSS and high SRT, where slow-growing microorganisms, such as nitrifying bacteria, are enriched. Thus, the MBR process, combined with anoxic denitrification, is able to achieve high nitrogen removal. However, the operation of the MBR with such higher SRT is not detrimental to biological phosphate removal, which is achieved by the discharge of phosphate containing biomass, and results in an excessive amount of phosphate in the cell from wastewater systems. A variety of MBR processes have been configured with biological nutrients removal (BNR) systems, with their performance demonstrated for different applications, but have only shown limited success. Therefore, it is common to see additional

chemical methods incorporated into the MBR process to satisfy regulatory requirements. The development of the MBR process coupled with more efficient BNR systems will be the primary topic over the next decade.

As mentioned above, conventional MBR processes utilizing particle separating MF/UF membranes are not able to meet the effluent quality for wastewater reclamation and reuse, and often require tertiary treatment, including tighter membrane technologies, such as RO. In the future, however, water supply and wastewater management systems will be more decentralized for the optimum use of energy and resources, and thus, complicated MBR systems with tertiary treatment should be further simplified and compact. Such technological demands have resulted in the development of a high rejection MBR process, where tighter membranes reject not only particles, but also dissolved micropollutants also. The NF-MBR is the first alternative MBR process, where an NF membrane is submerged in the bioreactor and produces much better effluent water qualities than those by conventional MF-MBR. NF membranes, with chemical stability as well as reasonable water productivity, need to be developed for improving their applicability in the future.

In addition to traditional pressure-driven tighter membranes, two different types of membrane have been utilized in high retention MBR processes; membrane distillation and forward osmosis membranes. The membrane distillation bioreactor (MDBR) integrates a conventional bioreactor with a thermally driven membrane distillation process, which uses microporous hydrophobic membranes for distillation. With warm wastewater, a modestly elevated temperature (30–80°C) can easily be obtained, and water vapor can be transported through the membrane and condense into water, but nonvolatile components, such as contaminants, are not able to pass through the membrane and thus, will be rejected by the membrane. The energy demand for such systems is known to be low compared to conventional MBR coupled with RO, although the economics of scale-up systems will need to be examined further. Biological environments specific to this process should be further understood for the better design and operation of the process in the future. Forward osmosis bioreactors (FO-MBR) are another recent development. The FO membrane process utilizing a concentration gradient as the driving

force is combined with a conventional bioreactor. Compared to RO processes, the FO process is characterized by low energy use and high recovery. A non-pressurized FO system is also known to enable increased fouling reversibility, thus making this process more suitable for wastewater applications where the fouling potential is very high. Although the benefits of such systems have been demonstrated in the literature, only limited performance data are available for small-scale applications. Therefore, newly developed high retention MBRs have a long way to go before they become a reality, but they are expected to attract many investigations and continue to grow as a promising technology to meet the demand for compact and reliable wastewater reuse systems.

Bibliography

Primary Literature

1. Bixio D, Thoeye C, Koning J, Joksimovic D, Savic D, Wintgens T, Melind T (2006) Wastewater reuse in Europe. *Desalination* 187:89–101
2. Bixio D, Thoeye C, Wintgens T, Ravazzini A, Miska V, Muston M, Chikurel H, Aharoni A, Joksimovic D, Melin T (2008) Water reclamation and reuse: implementation and management issues. *Desalination* 218:13–23
3. Cooper RC, Olivieri AW (1998) Infectious disease concerns in wastewater reuse. In: Asano T (ed) *Wastewater reclamation and reuse*. Technomic Publishing, Lancaster, pp 489–520
4. Dolnicar S, Schafer AI (2007) Desalination and recycling: Australians raise health, environment and cost concerns. *Desalination Water Reuse* 16:10–15
5. Dreizin Y (2006) Ashkelon seawater desalination project – off-taker's self costs, supplied water costs, total costs and benefits. *Desalination* 190:104–116
6. Feachem RG, Bradley DJ, Garelick H, Mara D (1983) *World sanitation and disease: health aspects of excreta and wastewater management by Bank Studies*, vol 3, Water supply sanitation. Wiley, Chichester
7. Fewtrell L, Bartram J (2001) *WHO guidelines, standards and health*. IWA Publishing, London
8. Force J (2010) Encouraging energy efficiency in US wastewater treatment. *Water21, Magazine of the IWA*, June 2009: 32–34
9. Hespánhol I, Prost AME (1994) WHO guidelines and national standards for reuse and water quality. *Water Res* 28(1): 119–124
10. Kamizoulis G (2008) Setting health based target for water reuse. *Desalination* 218:154–163
11. Kayaalp NM (1996) Regulatory framework in South Australia and reclaimed water reuse options and possibilities. *Desalination* 106(3):17–322

12. Kolodziejski J, Gasson C (2005) Water reuse markets 2005–2015: a global assessment and forecast. Media Analytics, Oxford (Global Water Intelligence)
13. Kretschmer N, Ribbe L, Gaese H (2003) Wastewater reuse for agriculture. *Technol Resour Manage Dev* 2:37–64
14. Levine AD, Asano T (2004) Recovering sustainable water from wastewater. *Environ Sci Technol* 38:201A–208A
15. Pearce GK (2008) UF/MF pre-treatment to RO in seawater and wastewater reuse applications: a comparison of energy costs. *Desalination* 222:66–73
16. Salgot M, Huertas E, Weber S, Dott W, Hollender J (2006) Wastewater reuse and risk: definition of key objectives. *Desalination* 187:29–40
17. Salgot M (2008) Water reclamation, recycling and reuse: implementation issues. *Desalination* 218:190–197
18. Reardon RD, Paranjape SV, Fousseureau XJ, DiGiano FA, Aitken MD, Kim JH, Chong S-Y, Cramer R (2005) Water for reuse: membrane treatment of 2ry effluent for subsequent use. WERF Final Report, North Carolina University
19. Toze S (2006) Water reuse and health risks – real vs. perceived. *Desalination* 187:41–51
20. USEPA (1992) Process design manual: guidelines for water reuse. Report No. EPA-625/R-92-004, Cincinnati
21. USEPA (2004) Guidelines for water reuse. Report No. EPA/625/R-04/108

Books and Reviews

- Asano T, Burton FL, Leverenz HL, Tsuchihashi R, Tchobanoglous G (2006) Water reuse: issues, technologies, and applications. Metcalf & Eddy/AECOM Press, McGraw-Hill, New York
- Asano T, Jimenez B (2008) Water reuse: an international survey of current practice issues and needs. IWA Publishing, London
- WHO (2006) A compendium of standards for wastewater reuse in the Eastern Mediterranean Region, WHO-EM/CEH/142/E

Wastewater Treatment and Control Through Wetlands

JAN VYMAZAL

Department of Landscape Ecology, Czech University of Life Sciences Prague, Kostelec nad Černými lesy, Prague, Czech Republic
ENKI, o.p.s., Třeboň, Czech Republic

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Types of Constructed Wetlands for Wastewater Treatment

The Use of Constructed Wetlands for Various Types of Wastewater

Future Directions

Bibliography

Glossary

Constructed wetland (CW) Constructed wetlands are engineered systems that have been designed and constructed to utilize the natural processes involving wetland vegetation, soils, and their associated microbial assemblages to assist in treating wastewater. Constructed wetlands offer several additional advantages compared to natural wetlands including site selection, flexibility in sizing, and most importantly, control over the hydraulic pathways and retention time.

CW with horizontal subsurface flow Constructed wetland where wastewater is fed in at the inlet and flows slowly through the porous medium under the surface of the bed in a more or less horizontal path until it reaches the outlet zone where it is collected before leaving via level control arrangement at the outlet. During this passage, the wastewater will come into contact with a network of aerobic, anoxic, and anaerobic zones.

CW with vertical flow Constructed wetlands where water is fed intermittently on the surface of the filtration bed and water percolates by gravity to the bottom where it is collected.

CW with free water surface Consists of basins or channels, with soil or another suitable medium to support rooted vegetation (if present) and water at a relatively shallow depth flowing through the unit.

Macrophytes Plants that are adapted to the life under anoxic and/or anaerobic soil conditions in wetlands.

Wetland Land that is saturated with water long enough to promote wetland or aquatic processes as indicated by poorly drained soils, hydrophytic vegetation, and various kinds of biological activity that are adapted to a wet environment.

Definition of the Subject and Its Importance

Wetlands are found on every continent except Antarctica and in every climate from the tropics to the frozen tundra. Any estimate of wetland extent in the world is difficult and depends on an accurate definition [1]. As the definitions vary, the estimated area of wetlands varies between 5.3 and 11.65 million square kilometers and between 3.5% and 7.7% [1–6]. As wetlands were considered as neither “true” terrestrial ecosystem nor “true” aquatic ecosystems, not many researchers were interested in wetland ecosystems. However, there has been an explosive growth of knowledge about, and a radical change of attitude toward wetlands since the 1950s [7]. Wetlands have been recognized as providing many ecosystem services such as hydrologic, biologic, biogeochemical, and cultural [8–12].

Natural wetlands have been used for wastewater treatment for centuries [13]. In many cases, however, the reasoning behind this use was disposal, rather than treatment and the wetland simply served as a convenient recipient that was closer than the nearest river or other waterway. Wetlands have been considered for a long time as “wastelands,” were scientifically neglected and, therefore, the impact of wastewaters on different wetlands was not properly assessed. Uncontrolled discharge of wastewater led in many cases to an irreversible degradation of many wetland areas [14]. Although significant improvement in the quality of the wastewater is generally observed as a result of flow through natural wetlands, the extent of their treatment capability is largely unknown [15]. Natural wetlands are still used for wastewater treatment under controlled conditions [16, 17], but the use of constructed wetlands has become more popular and effective around the world since the 1980s.

Constructed wetland treatment systems are engineered systems that have been designed and constructed to utilize the natural processes involving wetland vegetation, soils, and their associated microbial assemblages to assist in treating wastewater. They are designed to take an advantage of many of the same processes that occur in natural wetlands, but do so within a more controlled environment. Constructed wetlands are essentially wastewater treatment systems and are designed and operated as such, though many systems do support other functional values [18].

Constructed wetlands can be built with a much greater degree of control than natural systems, thus allowing the establishment of experimental treatment facilities with well-defined composition of substrate, type of vegetation, and flow pattern. In addition, constructed wetlands offer several additional advantages compared to natural wetlands including site selection, flexibility in sizing, and most importantly, control over the hydraulic pathways and retention time [15].

Introduction

In 1953, Dr. Käthe Seidel for the first time presented methods for improvement of inland waterways, which suffered from overfertilization, pollution from sewage, and siltation by means of appropriate plant species [19]. Between 1952 and 1956, Seidel carried out numerous experiments on the use of wetland plants for treatment of various types of wastewater, including phenol, dairy, or livestock wastewater [20–22]. In the early 1960s, Seidel intensified her trials to grow macrophytes in wastewater and sludge of different origin and she tried to improve the performance of rural and decentralized wastewater treatment, which was either septic tanks or pond systems with inefficient treatment. To overcome the anaerobic septic tank systems, she integrated a stage of primary sludge filtration in vertically percolated sandy soils planted with *Phragmites australis*. So the system consisted of an infiltration bed through which the sewage flowed vertically and an elimination bed with a horizontal flow [23]. This system was the basis for hybrid systems, which were revived at the end of the twentieth century. However, Seidel’s concept to apply macrophytes to sewage treatment was difficult to understand for sewage engineers who had eradicated any visible plants on a treatment site for more than 50 years [24] and therefore, it was no surprise that the first full-scale constructed wetlands were built outside Germany.

In spite of many prejudices among civil engineers about odor nuisance, attraction of flies, and poor performance in cold periods, the IJssel Lake Polder Authority in Flevoland in the Netherlands constructed its first free water surface constructed wetland (FWS CW) in 1967 [25]. Contrary to the North America and Australia, FWS CWs did not spread throughout Europe significantly while CWs with subsurface horizontal

(HF CWs) and vertical (VF CWs) flow drew much more attention at the end of the twentieth century [26]. However, FWS CWs are in operation in many European countries, for example, Sweden, Denmark, Poland, Estonia, or Belgium [27].

Types of Constructed Wetlands for Wastewater Treatment

Constructed wetlands could be classified according to the various design parameters but two most important criteria are water flow regime (surface and subsurface) and the type of macrophytic growth (Fig. 1). Different types of constructed wetlands may be combined with each other (so-called hybrid or combined systems) in order to exploit the specific advantages of the different systems. The quality of the final effluent from the systems improves with the complexity of the facility.

Constructed Wetlands with Free Water Surface

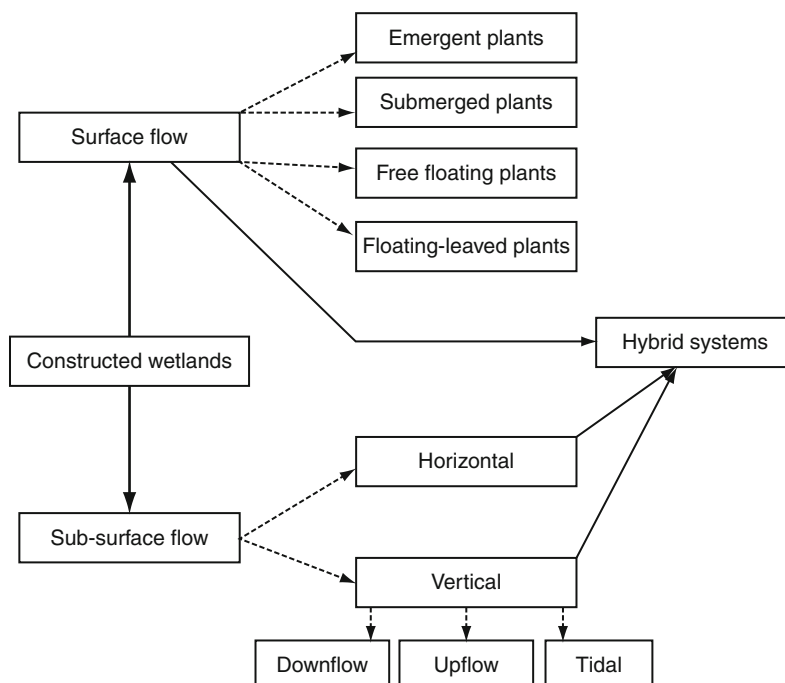
Constructed wetlands (CWs) with surface flow are commonly referred to as “free water surface” (FWS CWs).

They mimic the hydrologic regime of natural wetlands. In FWS CWs, treated wastewater enters one end of the basin, flows through an area with macrophytes, and is collected and discharged at the opposite end of the basin. The water is often shallow and moves through the wetland at a relatively slow velocity [28].

Systems with Free-Floating Macrophytes

Constructed wetlands with free-floating macrophytes (Fig. 2) consist of one or more shallow ponds in which plants float on the surface. The shallower depth and the presence of aquatic macrophytes in place of algae are the major differences between constructed wetlands with free-floating macrophytes and stabilization ponds [29]. Free-floating macrophytes are highly productive and belong to the fastest growing plants on the Earth. The most commonly used species in constructed wetlands are Water hyacinth (*Eichhornia crassipes*) and species of the family Lemnaceae (Duckweed) such as *Lemna minor*, *L. trisulca*, or *Spirodela polyrrhiza*.

A single Water hyacinth plant can produce approximately 65,000 offspring during a single season [30]



Wastewater Treatment and Control Through Wetlands. Figure 1

Types of constructed wetlands for wastewater treatment



Wastewater Treatment and Control Through Wetlands.
Figure 2

Constructed wetland with Water hyacinth (*Eichhornia crassipes*) in Yantian, South China as a part of hybrid constructed wetland for treatment of industrial wastewater (Photo: Jan Vymazal)

and ten plants can completely cover 0.4 ha of a water surface [31]. However, Water hyacinth does not survive the winter in temperate and cold climates, its growth rate is greatly reduced at temperatures below 10°C and, therefore, its use is restricted to subtropical and tropical regions. Moreover, the optimum water temperature is >25°C [32]. Duckweeds, compared to Water hyacinth, have a much wider geographic range as they are able to grow at temperatures as low as 1–3°C [15]. They even survive light freezes and continue to grow during the winter months but their growth optimum is at temperatures between 20°C and 30°C [33].

In CWs with free-floating macrophytes, organic substances are primarily removed by microbial degradation in water column by free-living bacteria or within a dense network of plant roots by attached bacteria. The degradation proceeds under both aerobic and anoxic/anaerobic conditions but aerobic processes are usually restricted only to the close vicinity of the roots

or to the initial stages of plant development. As the plant density increases, oxygen diffusion into the water is restricted and also the plants cover completely the water surface and prevent oxygen-producing algae from growing in the water column [34]. Suspended solids are removed mainly through gravity sedimentation in the zone under the surface layer of floating plants but in Water hyacinth-based systems, substantial amount of suspended solids is retained in the root zone. The sedimentation rate is usually high as the plant layer on the surface eliminates the water movement and turbulence caused by wind. Phosphorus removal is mostly the result of the plant uptake and therefore it is desirable to keep plants in the exponential growth phase. This is realized by frequent harvesting of the plant biomass [35–37]. The harvesting has another beneficial effect because it opens the water surface for light penetration into the water and surface re-aeration. Nitrogen is removed via several processes including plant uptake (and subsequent harvesting), ammonia volatilization, and microbial nitrification/denitrification [38]. Ammonia volatilization is limited to early stages of wetland development when algae grow in water column and as a consequence of photosynthesis the pH values could be high enough for NH_3 volatilization. Nitrification occurs within the aerobic root zone and denitrification occurs in the much larger reduced zones in the water column. However, the nitrification/denitrification complex is usually not very effective for nitrogen removal because of limited nitrification in the system [38].

The CWs with Water hyacinth are very effective in pollutant removal. Vymazal [39] summarized the treatment efficiency of 14 systems in the USA and Japan and the average BOD_5 , suspended solids, TN, and TP removal efficiencies amounted to 76%, 64%, 60%, and 47%, respectively. The mean outflow concentrations of respective parameters were 14.1, 9.3, 6.6, and 2.2 mg L^{-1} .

During the 1970s and early 1980s, constructed wetlands with free-floating macrophytes were intensively studied but the operational needs associated with large-scale applications have never been developed [29, 40]. Probably, the major obstacles to wider application of CWs with free-floating macrophytes are very high maintenance and operation costs and high amount of the biomass with no other use [41].

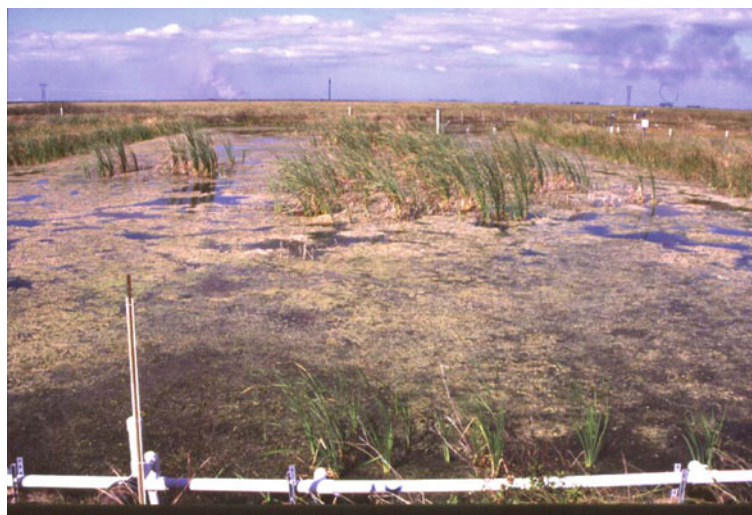
Systems with Submerged Macrophytes Submerged aquatic macrophytes have their photosynthetic tissue entirely submerged. Various experiments have proven that minerals can be taken up directly by shoot tissues of submerged plants; however, there is also no question regarding the uptake capability of nutrients by the roots of these plants. Most studies confirmed that although shoot uptake occurs, the major source for nutrients are sediments [42, 43]. Submerged plants, however, only grow well in oxygenated water and therefore cannot be used in wastewater with a high content of readily biodegradable organic matter because the microbial decomposition of the organic matter will create anoxic conditions [15]. In addition, the turbidity of the water must not be too high to prevent light transmission to the plants to support their photosynthetic activity [38]. Because of limiting conditions mentioned above (low light, high turbidity), the prime potential use of submerged macrophyte-based wastewater treatment systems (Fig. 3) is therefore for “polishing” secondary treated wastewaters [15].

There are many submerged species but the most promising are probably *Egeria densa* (Brazilian elodea), *Elodea canadensis* (Canadian waterweed), *Elodea nuttallii* (Western waterweed), *Ceratophyllum demersum* (Coontail), *Hydrilla verticillata* (Hydrilla),

Cabomba caroliniana (Fanwort), *Myriophyllum* spp. (Watermilfoil), and *Najas guadalupensis* (Southern naiad). Submerged macrophytes are very often covered with periphyton, an attached community usually dominated by algae. The periphyton has a beneficial effect on nutrient removal but it also absorbs photosynthetically active radiation before it reaches the leaf surface of vascular plants thus making submerged plants light-limited [44]. The design criteria for CWs with submerged macrophytes still have not been fully developed, but there have been some good examples reported in the literature [45–47].

Constructed Wetlands with Floating-Leaved Macrophytes

Floating-leaved macrophytes include plant species that are rooted in the substrate and their leaves on long peduncles float on the water surface. Typical examples are Water lilies (*Nymphaea* spp.), Spatterdock (*Nuphar lutea*), or Indian lotus (*Nelumbo nucifera*). In fact, this type combines the features of constructed wetlands with other types of macrophytes. The plants have huge underground rhizomes, leaf peduncles grow through the water column, and leaves float on the water surface. However, peduncle density is much lower as compared to submerged or emergent species and leaves rarely cover the entire surface of the



Wastewater Treatment and Control Through Wetlands. Figure 3

Constructed wetland with submerged vegetation (*Najas guadalupensis*) in the Florida Everglades Nutrient Removal Area (Photo: Jan Vymazal)

pond. So far, only a few constructed wetlands have used these plants [48] and therefore any sound conclusions on their efficiency and operation could be made.

Constructed Wetlands with Emergent Macrophytes

A typical free water surface constructed wetland (FWS CW) with emergent macrophytes is a shallow sealed basin or sequence of basins, containing 20–30 cm of rooting soil, with a water depth of 20–60 cm. Dense emergent vegetation covers significant fraction of the surface, usually more than 50% (Fig. 4). Besides planted macrophytes, naturally occurring species may be present [49]. Areas of open water may be incorporated in design to improve hydraulics and enhance wildlife habitat [28]. One of the primary design purposes is to provide wastewater with reactive biological surfaces. For FWS CWs, many plant species are used but the most common species are *Phragmites australis* (Common reed), *Typha* spp. (Cattails), and *Scirpus* (*Schoenoplectus*) spp. (Bulrush).

Suspended solids are mostly removed by gravitational settling. The largest and heaviest particles will predominantly settle out in the inlet open water zone, while slightly smaller and lighter particles may only settle out after flowing into wetland vegetation.

Wetland vegetation promotes this enhanced sedimentation by reducing water column mixing and resuspension of particles from the sediment surface. The smallest of particles such as clay colloids may be removed by adhesion onto surfaces within the water column. The main surfaces within the water column are the emergent macrophytes and the biofilms growing on the surface of these plants [50].

Settleable organics are rapidly removed in FWS systems with emergent vegetation under quiescent conditions by deposition and filtration. Attached and suspended microbial growth is responsible for the removal of soluble organic compounds, which are degraded aerobically in the water column as well as anaerobically near the bottom within the litter layer. Oxygen is supplied to the wetland water column by diffusion through the air–water interface and via the photosynthesis of plants in the water column, namely algae [29, 50].

Nitrogen is most effectively removed by volatilization and nitrification/denitrification. Volatilization is very likely under higher pH values caused by photosynthesis of both planktonic and attached algae. Nitrification proceeds in aerobic zones, while nitrate is converted to free nitrogen or nitrous oxide in the



Wastewater Treatment and Control Through Wetlands. Figure 4

FWS CW Mc Grath Hill, Hawkesbury near Sydney, Australia planted with a mixture of *Phragmites australis* (Common reed) and *Bolboschoenus fluviatilis* (River bulrush) (Photo: Jan Vymazal)

anoxic/anaerobic zones by denitrifying bacteria. Aerobic zones are usually in the water column and especially close to the water surface, while anaerobic zones are mostly found in the litter layer on the bottom. FWS CWs with emergent vegetation provide sustainable removal of phosphorus, but at relatively slow rates. Phosphorus removal occurs from adsorption, absorption, complexation, and precipitation. However, the major retention process – precipitation with Al, Fe, and Ca ions – is limited by little contact between the wastewater and the soil [29, 51]. Substantial amounts of nitrogen and phosphorus may be stored in the peat/litter compartment. Algal and microbial uptake may be high but this retention is a short-term process and nutrients are washed out from the detritus back to the water [52].

Removal of nutrients via plant uptake is usually low as compared to inflow load. In addition, if the biomass is not harvested nutrients are released back into water during the biomass decay. Some of these nutrients could be subject to long-term accumulation in the soil, hence contributing to the wastewater-treatment capacity of the system [53].

Wetlands provide a suitable combination of physical, chemical, and biological factors for the removal of pathogenic organisms. Physical factors include mechanical filtration, exposure to ultraviolet radiation, and sedimentation. Chemical factors include oxidation, exposure to biocides excreted by some wetland plants, and absorption by organic matter. Biological removal mechanisms include antibiosis, predation by nematodes and protists, attack by lytic bacteria and viruses, and natural die-off [54].

In Europe, this technology started during the late 1960s. One of the first examples of full-scale FWS CWs with emergent vegetation are those built in 1967 in Lelystad, the Netherlands [25] and near Keszthely, Hungary [55]. This type of constructed wetlands have not spread too much throughout Europe as compared to subsurface constructed wetlands but recently many wetlands have been created in several countries, especially in Sweden and Denmark. More than 2,350 ha of wetlands have been created in Sweden in the agricultural landscape between 1996 and 2002 with different kinds of subsidies and in Denmark about 3,200 ha have been created until 2004 with the primary aim to reduce nitrogen input into Baltic Sea [27].

In North America, the FWS technology started with the ecological engineering of natural wetlands for wastewater treatment in the early 1970s in Florida [56] and Michigan [57]. Industrial stormwaters and process waters were applied to FWS CWs in North Dakota in 1975 [58] and in the early 1980s the use of FWS constructed wetlands for urban stormwater treatment was pioneered in California [59] and for mine wastewater throughout the USA [60]. At present, there are hundreds, and probably thousands of FWS constructed wetlands with emergent macrophytes treating municipal and industrial wastewaters, agricultural wastes and runoff, mine drainage waters, and stormwaters in North America [61]. Free water surface CWs are also commonly used in Australia and New Zealand, especially for treatment of municipal wastewater, stormwater, and pasture runoff [62, 63].

Constructed wetlands with free water surface and emergent vegetation are very effective in removing suspended solids, organics, and microbial pollution. Removal of nitrogen and phosphorus is moderate.

Constructed Wetlands with Floating Mats of Emergent Macrophytes Some emergent macrophytes are capable of forming floating mats, even though their individual plants are not capable of such existence (Fig. 5). Under field conditions, the floating islands of emergent macrophytes may occur naturally as a consequence of bottom disturbance. *Typha* spp. (cattails), *Glyceria maxima* (Mannagrass), *Phragmites australis* (Common reed), *Cyperus papyrus* (Papyrus), *Hydrocotyle umbellata* (Pennywort), or *Alternanthera philoxeroides* (Alligator weed) are examples of plants that are capable of growing in mats.

The floating mats need the support structures, which keep plants in upright position. This is realized either by various netlike structures or by floating, usually styrofoam islands. This technology spread gradually around the world [64, 65] but any design guidelines have not been developed yet.

Constructed Wetlands with Subsurface Flow

Constructed wetlands with subsurface flow may be classified according to the direction of flow into horizontal (HF or HSF) and vertical (VF). Vertical-flow treatment wetlands could be further categorized into



Wastewater Treatment and Control Through Wetlands. Figure 5

FWS CW with floating mats of emergent macrophytes for treatment of stormwater runoff (Bornem, Belgium). Plants are kept in position with coconut strings (Photo: Jan Vymazal)

downflow and upflow whether the wastewater is fed onto the surface or to the bottom of the wetland (Fig. 1). In these systems, there is no free water surface and processes, which are responsible for pollution removal occur in the layer of filtration material.

Systems with Horizontal Subsurface Flow (HF CWs)

In HF CWs, mechanically pretreated wastewater flows slowly under the surface of the filtration bed, which is filled with porous material planted with emergent macrophytes (Fig. 6). During this passage, the wastewater will come into contact with a network of aerobic, anoxic, and anaerobic zones. The aerobic zones occur around roots and rhizomes that leak oxygen into the substrate [66, 67]. The filtration bed is sealed from the surrounding by an impermeable layer, in most cases plastic liner (PVC, DHPE, LDPE) to prevent leakage to the groundwater (Fig. 6). In order to protect the liner from damage, a geotextile is commonly used (Fig. 6).

The technology of wastewater treatment by means of constructed wetlands with horizontal subsurface flow was started in Germany based on research by K. Seidel commencing in the 1960s [20, 22, 23] and by R. Kickuth in the 1970s [68, 69]. The first full-scale HF Kickuth's type CW for treatment of municipal

sewage was put in operation in 1974 in the community Liebenburg-Othfresen [69]. In 1981, the first HF constructed wetland was built in Weinitzen near Graz, Austria [70] and in 1983, the HF CW technology was introduced in Denmark where about 80 HF CWs were built by 1987. Despite problems with surface flow, soil-based systems exhibited high treatment effect for organics and suspended solids if vegetated bed area $3\text{--}5\text{ m}^2\text{ PE}^{-1}$ was used (PE = population equivalent, $60\text{ g BOD}_5\text{ day}^{-1}$) [66].

In 1985, first two HF CWs were built in the UK (here called Reed Bed Treatment Systems) and by the end of 1986 more than 20 HF CWs were designed [71]. The major change in the design was the use of very coarse filtration material (5–10 mm fraction), which ensured subsurface flow similarly to former design by Seidel [23]. Also, specific area of $5\text{ m}^2\text{ PE}^{-1}$ was used in the UK. During the 1990s, HF CWs were built in many European countries and HF CWs became the most widely used concept in Europe in the 1990s [26, 72].

In the USA, between 1972 and 1976, F. Spangler and colleagues started the research aimed at the use of constructed wetlands with horizontal subsurface flow in Seymour, Wisconsin to treat municipal wastewater [73, 74]. This study was followed by other experiments



Wastewater Treatment and Control Through Wetlands. Figure 6

Constructed wetland with horizontal subsurface flow. *Top* – plastic liner (*right*) covered with a geotextile (*left*). *Bottom* – filtration bed filled with crushed rock (4–8 mm) with inflow and outflow zones filled with stones (32–64 mm) (*right*) planted with Reed canarygrass (*Phalaris arundinacea*) (*left*) (Photos: Jan Vymazal)

in New York [75] and Mississippi [76] and during the early 1980s in California [77] and Kentucky [78]. North America was slower to adopt subsurface technology as compared to Europe. However, in recent years, the use of these systems has drawn more attention and it is estimated that there are about 8,000 subsurface constructed wetlands at present [61]. HF CWs in Australia were also adopted in the late 1970s and early 1980s [79, 80] but at present mostly small on-site systems in rural areas are used treatment of domestic sewage [81].

Since the mid-1980s, the concept of HF CWs constructed wetlands has been used in South Africa [82]. At the end of the twentieth century, constructed wetland became more popular in tropical parts of Africa and there are now many fine examples of all types of constructed wetlands including HF systems treating municipal sewage as well as industrial wastewaters and mine drainage waters in [83, 84]. In Asia, the first information about the use of HF CWs appeared in the literature only in the early 1990s from India and China [85, 86] but at present, HF CWs are in

operation in many other Asian countries such as Korea, Taiwan, Japan, Nepal, Malaysia, or Thailand (e.g., [87, 88]). In South and Central America, the use of HF CWS is not widely distributed but this technology has recently got more attention (e.g., [89, 90]).

The early HF CWs were filled with local soil in order to provide high filtration effect. However, this filtration material did not provide sufficient hydraulic conductivity to keep the flow under the surface of filtration beds. After a short period of operation, the surface flow occurred. According to Kickuth's theory, the root formation would increase the hydraulic conductivity of the soil matrix by about two orders of magnitude within 2–4 years of operation [91], but this phenomenon has never been proved (e.g., [70]) and based on recent studies, the presence of plant roots and rhizomes in the HF beds will have a negative effect on hydraulic conductivity (e.g., [92]). However, most of the hydraulic conductivity decline is apparently associated with sediments trapped within the filtration bed from the wastewater. Therefore, more coarse filtration material were required and it turned out that gravel or crushed rock provide required hydraulic conductivity and do not clog rapidly when the wastewater is properly mechanically pretreated [71]. Since the mid-1980s, washed gravel or crushed rock with the grain size between 5 and 20 mm is most commonly used. In

order to keep the water below the filtration bed surface, it is also important to have a good mechanical pretreatment and reliable outflow structure. The small secondary systems usually use septic tanks, larger secondary systems most commonly use a combination of screens, grit chamber, and Imhoff tank. However, many other combinations are possible.

For good hydraulics of HF CWs also distribution and collection of water is important. In HF constructed wetlands, the aim is to get even distribution across the full-sectional area of the inlet end of the bed. The early systems used long channels with castellated or vee-notch weirs, but this distribution system had problems with screenable material collected at edges and causing maldistribution and therefore they were replaced with single pipes with adjustable tees or orifices spread along their length and later with perforated pipes either buried or laid on the surface of the distribution zone. At the outlet end, most systems usually have a perforated agricultural drain pipe buried on the bottom of the outflow zone filled with stones. The pipe is connected with a sump where the water level is controlled by socketed pipes, swiveling elbows, or flexible hose(s) (Fig. 7), which can be held in position by a chain or rope.

The macrophytes growing in constructed wetlands have several properties in relation to the treatment



Wastewater Treatment and Control Through Wetlands. Figure 7

Swiveling elbow hanging on a chain in HF CW Sklené, Czech Republic (Photo: Jan Vymazal)

process that make them an essential component of the design. In HF CWs plants, macrophytes (1) insulate the surface during the periods of cold weather, (2) provide substrate for attached bacteria and oxygenation (even very limited) of areas adjacent to roots and rhizomes, (3) release root exudates that may have antimicrobial properties, and (4) sequester nutrients from the wastewater in the (aboveground) biomass that could be removed via harvesting [93]. The most commonly used plant around the world is *Phragmites australis* (Common reed) but many other species such as *Phalaris arundinacea* (Reed canarygrass), *Scirpus* (*Schoenoplectus*) spp. (Bulrush), *Iris pseudacorus* (Yellow flag) are used including the local species.

HF CWs provide high removal of suspended solids, organics, and microbial pollution; removal of nitrogen is moderate and removal of phosphorus is low unless special filtration media with high sorption capacity are used. Organic compounds are degraded by aerobic as well as anaerobic microorganisms attached to the plant's underground organs (i.e., roots and rhizomes) and media surface. The aerobic processes are usually restricted only to the close vicinity of roots where oxygen leaks to the rhizosphere as diffusion of oxygen from the atmosphere to a saturated filtration bed is very slow [94] and therefore anoxic and anaerobic decomposition play an important role in HF constructed wetlands.

The major removal mechanism for nitrogen in HF CWs is microbial nitrification/denitrification. However, the field measurements have shown that the oxygenation of the rhizosphere of HF constructed wetlands is insufficient and, therefore, the incomplete nitrification is the major cause of limited nitrogen removal. Volatilization, adsorption, and plant uptake play much less important role in nitrogen removal in HF constructed wetlands. Volatilization is limited by the fact that HF CWs do not have free water surface and algal activity is negligible in these systems. Nitrogen removal via plant uptake and subsequent harvesting is usually low as compared to inflow loads [95].

Phosphorus is removed primarily by ligand exchange reactions, where phosphate displaces water or hydroxyls from the surface of Fe and Al hydrous oxides. However, media used for HF wetlands (e.g., pea gravel, crushed stones) usually do not contain great quantities of Fe, Al, or Ca and therefore, removal of

phosphorus is generally low [95]. Recently, several filtration materials such as LECA (lightweight clay aggregates) have been tested in constructed wetlands [96]. The removal of phosphorus is very high but it is important to realize that sorption and precipitation are saturable and the sorption decreases over the time. It has been found that removal of phosphorus through plant harvesting is negligible and forms only a tiny fraction of the amount removed. However, this mechanism may play more significant role in nutrient removal in tropical and partially subtropical regions where the growth seasonality and nutrient translocations between above- and below-ground parts are minimal [95].

Microbial pollution removal is mainly achieved through a combination of physical, chemical, and biological factors [97]. Physical factors include mainly filtration and adsorption. Chemical and biological factors may include oxidation, exposure to biocides excreted by a number of wetland plants, antimicrobial activity of root exudates, predation by nematodes and protists, activity of lytic bacteria or viruses, retention in biofilms and natural die-off [21, 93, 98].

Constructed wetlands with horizontal subsurface flow are primarily used to remove organics and suspended solids from the various types of wastewater. They are not effective in ammonia removal due to lack of oxygen in filtration bed but they provide suitable conditions for denitrification. Also, phosphorus removal is effective only when special filtration materials are used.

Systems with Vertical Subsurface Flow (VF CWs)

The earliest form of vertical-flow (VF) system is that of Seidel in Germany [23], which was used to oxygenate anaerobic septic effluents. Similar systems in the Netherlands were called "infiltration fields" [99]. Interest in the particular process seemed to wane but it was revived in the last decade because of the need to provide low ammonia concentrations in the discharged water.

Vertical-flow (VF) constructed wetlands comprise a sandy filtration bed planted with macrophytes. The wastewater is fed intermittently on the wetland surface (Fig. 8) and the next batch of wastewater is flushed only after the previous batch percolated through the filter. Wastewater is collected by a drainage network at the base. This feeding mode drains the bed completely free and it allows air to refill the bed. As a result,



Wastewater Treatment and Control Through Wetlands. Figure 8

Wastewater distribution in an experimental VF CW Carrion de Cespedes, Seville, Spain (picture taken before planting)
(Photo: Jan Vymazal)

vertical-flow CWs provide good oxygen transfer and hence the ability to nitrify. The major purpose of macrophyte presence in VF CWs is to help maintain the hydraulic conductivity of the bed. By far, the most frequently used species is *Phragmites australis*. The early vertical-flow systems were usually staged systems with parallel and series beds. Such VF systems are now called first-generation VF systems [100]. Recently, VF systems with only one bed have been used. These systems are called second-generation VF CWs or compact vertical-flow beds [101].

The removal mechanisms in VF CWs are the same as in HF CWs. However, the major difference is the extent of processes responsible for nitrogen transformation and removal. VF CWs provide much higher degree of nitrification but removal of total nitrogen is comparable (Table 1) because the redox conditions in VF CWs are not suitable for denitrification and nitrogen is discharged as nitrate [95].

Vertical-flow constructed wetlands are very often used for on-site treatment [100, 101] but some countries such as France, the VF CWs are used to treat municipal sewage [102]. VF systems are getting more popular especially for their ability to remove ammonia. The most commonly used format is that a series of first

stage beds dosed in rotation acts as combination of a sludge drying reed bed and pretreatment and the second stage is a “conventional” VF bed [102].

Besides vertical-flow CWs fed on the surface, there are VF wetlands that are fed on the bottom of the filter bed (so-called “upflow” systems). The water percolates upward and then it is collected either near the surface or on the surface of the wetland bed. These systems have commonly been used in Brazil since the 1980s [103], but they are used now in limited numbers in other countries as well.

Tidal flow systems are a new form of VF system [100]. These systems were developed to try to overcome some of the problems that were seen in the early forms of VF systems related to clogging of the surface. Upflow systems have been used for about 20 years but they suffer from the problem that distribution is below the surface and hence hidden from the observers. In tidal flow systems at the start of the treatment cycle the wastewater is fed to the bottom of the bed into the aeration pipes. It then percolates upward until the surface is flooded. When the surface is completely flooded the pump is then shut off, the wastewater is then held in the bed in contact with the microorganisms growing on the media. A set time later the wastewater is drained downward and

Wastewater Treatment and Control Through Wetlands. Table 1 Treatment performance of downflow, vertical-flow, and horizontal-flow constructed wetlands: worldwide experience

	Vertical flow				Horizontal flow			
	In	Out	Eff. (%)	<i>n</i>	In	Out	Eff. (%)	<i>n</i>
BOD ₅	309	21	87.9	97	170	42	73.4	438
COD	547	70	78.6	115	427	143	62.7	292
TSS	188	18	77.1	74	141	35	68.6	367
TP	10.6	4.6	48.3	94	9.6	4.8	42.8	272
TN	70	37.6	44.0	64	63.1	36	39.7	208
NH ₄ –N	56.4	10.6	78.9	94	36.1	22.1	22.6	305
NO ₃ –N	0.62	25.3		70	8.0	5.9	26.0	210
FC*	5.95	3.0	97.6	20	8.22	5.17	92.4	95

In = inflow, Out = outflow, Eff. = efficiency, *n* = number of systems. *FC = fecal coliforms (log CFU 100 ml⁻¹) (Data from Vymazal and Kröpfelová 2008)

after the water has drained from the bed the treatment cycle is complete and air diffuses into the voids in the bed [100].

Hybrid Constructed Wetlands

Various types of constructed wetlands may be combined in order to achieve higher treatment effect, especially for nitrogen. Many of these systems are derived from original hybrid systems developed by Seidel at the Max Planck Institute in Krefeld, Germany. The process is known as the Seidel system, the Krefeld system, or the Max Planck Institute Process (MPIP) [23]. The design consisted of two stages of several parallel VF beds followed by two or three HF beds in series. This configuration provides nitrification in the first (VF) wetland and denitrification in the second (HF) wetland. Such systems were built, among others, in France [104], the UK [105], Norway [106], or Belgium [107].

In the mid-1990s, Johansen and Brix [108] introduced a HF-VF hybrid system with a large HF bed placed first and a small VF bed as the second stage. In this system, nitrification takes place in the vertical-flow stage at the end of the process sequence. If nitrate removal is needed it is then necessary to recirculate the effluent back to the front end of the system where denitrification can take place. It is possible to recycle the wastewater to the pretreatment unit (e.g., septic tank) or to the HF CW

where conditions are favorable for denitrification (anoxic/anaerobic conditions and sufficient concentration of organics necessary for denitrification) [109].

Recently, hybrid constructed wetlands have combined even more types of CWs including FWS systems in order to achieve better treatment performance of the system. The FWS CWs help to oxygenate the water but very often contribute to the increase of suspended solids due to the presence of planktonic algae in the outflow. For example, Masi et al. [110], Mander et al. [111], and Brix et al. [112] reported the use of VF-HF-FWS-pond hybrid system to treat winery wastewaters in Ornellaia (Italy), sewage in Kõo (Estonia), and sewage in tsunami-destroyed area at the island of Phi Phi Don (Thailand).

The Use of Constructed Wetlands for Various Types of Wastewater

Constructed wetlands have mostly been used to treat municipal or domestic sewage. However, over the period of 4 decades, the constructed wetlands have also been used to treat nearly all types of wastewaters including industrial and agricultural, landfill leachate, mine drainage waters, and runoff waters (Table 2).

When municipal sewage treatment is concerned besides the common pollutants such as organics, suspended solids and nutrients also specific compounds,

such as linear alkylbenzene sulfonates, the most widely used synthetic surfactants [115], or PPCP (Pharmaceuticals and Personal Care Products) [116] are monitored. For both groups of compounds, constructed wetlands proved to be a suitable treatment technology.

Constructed wetlands are used to treat a wide variety of industrial wastewaters (Table 2). Industrial wastewaters vary widely in composition and also strength and some of these wastewaters such as those from food processing, distillery, winery or abattoir belong to the most concentrated wastewaters especially because of high concentrations of organics. However, it has been shown that constructed wetlands (Fig. 9) can handle these wastewaters successfully [117]. Also wastewaters from agroindustry operations are very strong and very often contain very high concentrations of suspended solids, organics, and ammonia. Therefore, it is necessary to include several steps of pretreatment and constructed wetlands are usually used as final treatment step.

Constructed wetlands are commonly used for treatment of various kinds of runoff. They improve water quality, modify the flow rates by storing water temporarily, attenuate flow, and reduce downstream scouring and erosion. Target pollutants differ among types of runoff. In highway, road and urban (Fig. 10) runoff, the

targets are suspended solids, organics, oil and grease, and heavy metals [118]. In nursery, greenhouse, agricultural, pasture, and golfcourse (Fig. 11) runoff, the major targets are nitrates, phosphorus, and herbicides [119]. In airport runoff, the major targets are de-icing and anti-icing compounds applied to the aircraft and taxiways [120].

Wastewater Treatment and Control Through Wetlands.
Table 2 The use of constructed wetlands for various types of wastewater [28, 29, 114, 115]

Industrial	Petrochemical, pulp and paper, electric utility, abattoir, tool industry, power station, sugar industry, olive mill, potato processing, soft drinks, explosives, chemicals, cheese, herbicides, textile, tannery, distillery, winery, laundry, fish farm, aquaculture (fish, shrimp)
Agriculture	Swine, livestock, poultry, dairy
Runoff	Dairy pasture, tile drainage, agriculture (row crops), residential areas, roads, highways, airport, golf-courses, greenhouse, nursery
Mine waters	Acid coal acid lignite, alkaline coal, coal slurry, metals (Cu, Zn, Pb, U), spoil heaps
Leachate	Landfill, compost, mushroom farm, wood



Wastewater Treatment and Control Through Wetlands. Figure 9
HF CW planted with *Typha domingensis* for treatment of abattoir wastewater in San Jacinto, Uruguay (Photo: Jan Vymazal)



Wastewater Treatment and Control Through Wetlands. Figure 10

FWS CW for urban stormwater runoff treatment in Woodcroft Estate near Sydney, Australia (Photo: Jan Vymazal)



Wastewater Treatment and Control Through Wetlands. Figure 11

FWS CW for treatment of golfcourse runoff (Rivertowne Country Club, Charleston, South Carolina, USA) (Photo: Jan Vymazal)

A very large application area for constructed wetlands is the treatment of mine drainage waters from coal (Fig. 12) and metal ores mining activities (Table 2). The target parameters are sulfates, iron, manganese,

aluminum, acidity and alkalinity in coal mining drainage, and heavy metals in metal ores drainage waters.

Constructed wetlands are also used to treat various leachates. The most systems have been built to treat



Wastewater Treatment and Control Through Wetlands. Figure 12

Constructed wetland with free water surface and emergent vegetation (*Typha latifolia*) in Monastery Run, Pennsylvania, USA for treatment of mine drainage waters (Photo: Jan Vymazal)

landfill leachates [121, 122]. This application is getting more attention nowadays and examples of constructed wetlands could be found in Canada, the USA, Slovenia, Norway, Poland, the UK, among others [114].

Future Directions

Constructed wetlands have evolved over the period of nearly 4 decades into a widely accepted alternative for wastewater treatment of all types of wastewaters, despite the mistrust of many civil engineers and water authorities. These systems are very effective in removing pollution from wastewater and at the same time their operation and maintenance costs are much lower as compared to conventional treatment systems. Various types of constructed wetlands provide specific advantages and therefore their combination may enhance the treatment effect.

Indeed, constructed wetlands do have their “weak points” such as phosphorus removal. So far, constructed wetlands have not been recommended as an appropriate technology for phosphorus removal unless specific filtration materials are used. However, the artificially prepared materials are usually expensive and therefore intensive efforts are made to test various waste materials such as

blast and electric arc furnace steel slags or fly ash (inorganic waste product from coal or oil shale combustion) to enhance phosphorus removal [123, 124].

It is desirable to pay more attention to secondary benefits of constructed wetlands designed primarily for wastewater treatment. For examples, the ecosystem services on the landscape are often neglected and seldom studied. Also the use of plant biomass from constructed wetlands is only scarce [125].

Bibliography

Primary Literature

1. Mitsch WJ, Mitsch RH, Turner RE (1994) Wetlands of the old and new worlds: ecology and management. In: Mitsch WJ (ed) Global wetlands: old world and new. Elsevier Science, Amsterdam, pp 3–56
2. Maltby E, Turner GE (1983) Wetlands of the world. Geogr Mag 55:12–17
3. Matthews E, Fung I (1987) Methane emission from natural wetlands: global distribution, area, and environmental characteristics of sources. Glob Biogeochem Cycles 1:61–86
4. Aselman I, Crutzen PJ (1989) Global distribution of natural freshwater wetlands and rice paddies, their net primary productivity, seasonality and possible methane emissions. J Atmos Chem 8:307–358

5. Aselman I, Crutzen PJ (1990) A global inventory of wetland distribution and seasonality, net primary productivity, and estimated methane emissions. In: Bouwman AF (ed) *Soils and the greenhouse effect*. Wiley, New York, pp 441–449
6. Patten BC (1990) Introduction and overview. In: Patten BC (ed) *Wetlands and shallow continental water bodies*, vol 1. SPB Academic, The Hague, pp 3–8
7. Williams M (1990) Understanding wetlands. In: Williams M (ed) *Wetlands: a threatened landscape*. Basil Blackwell, Oxford, pp 1–41
8. Chen ZM, Chen GQ, Chen B, Zhou JB, Yang ZF, Zhou Y (2009) Net ecosystem services value of wetlands: environmental economic account. *Commun Nonlinear Sci Numer Simul* 14:2837–2843
9. Geber U, Björklund J (2002) The relationship between ecosystem services and purchased input in Swedish wastewater treatment systems – a case study. *Ecol Eng* 19:97–117
10. Moreno D, Pedrocchi C, Comin FA, García M, Cabezas A (2007) Creating wetlands for the improvement of water quality and landscape restoration in semi-arid zones degraded by intensive agricultural use. *Ecol Eng* 30:103–111
11. Löffler H (1990) Human uses. In: Patten BC (ed) *Wetlands and shallow continental water bodies*, vol 1. SPB Academic, The Hague, pp 17–27
12. Sather JH, Smith RD, Larson JS (1990) Natural values of wetlands. In: Patten BC (ed) *Wetlands and shallow continental water bodies*, vol 1. SPB Academic, The Hague, pp 373–387
13. Stainbridge HH (1976) History of sewage treatment in Britain, part 5. Land treatment. Institute of Water Pollution Control, Maidstone
14. Wentz WA (1987) Ecological/environmental perspectives on the use of wetlands in water treatment. In: Reddy KR, Smith WH (eds) *Aquatic plants for water treatment and resource recovery*. Magnolia, Orlando, pp 17–25
15. Brix H (1993) Wastewater treatment in constructed wetlands: system design, removal processes, and treatment performance. In: Moshiri AG (ed) *Constructed wetlands for water quality improvement*. CRC Press, Boca Raton, pp 9–22
16. Kadlec RH (2009) The Houghton Lake wetland treatment project. *Ecol Eng* 35:1285–1286
17. Mander Ü, Jenssen P (eds) (2002) *Natural wetlands for wastewater treatment in cold climates*. WIT Press, Southampton
18. Hammer DA, Bastian RK (1989) Wetland ecosystems: natural water purifiers? In: Hammer DA (ed) *Constructed wetlands for wastewater treatment*. Lewis, Chelsea, pp 5–19
19. Seidel K (1953) Pflanzungen zwischen Gewässern und Land. *J Max Planck Inst* 17–20
20. Seidel K (1965) Phenol-Abbau in Wasser durch *Scirpus lacustris* L. während einer versuchsdauer von 31 Monaten. *Naturwissenschaften* 52:398–406
21. Seidel K (1976) Macrophytes and water purification. In: Tourbier J, Pierson RW (eds) *Biological control of water pollution*. Pennsylvania University Press, Philadelphia, pp 109–122
22. Seidel K (1961) Zur Problematik der Keim- und Pflanzengewässer. *Verh Internat Verein Limnol* 14:1035–1039
23. Seidel K (1965) Neue Wege zur Grundwasseranreicherung in Krefeld, Vol. II. Hydrobotanische Reinigungsmethode. *GWF Wasser/Abwasser* 30:831–833
24. Börner T, von Felde K, Gschlössl T, Gschlössl E, Kunst S, Wissing FW (1998) Germany. In: Vymazal J, Brix H, Cooper PF, Green MB, Haberl R (eds) *Constructed wetlands for wastewater treatment in Europe*. Backhuys, Leiden, pp 169–190
25. De Jong J (1976) The purification of wastewater with the aid of rush or reed ponds. In: Tourbier J, Pierson RW (eds) *Biological control of water pollution*. Pennsylvania University Press, Philadelphia, pp 133–139
26. Vymazal J (2005) Constructed wetlands with horizontal sub-surface flow and hybrid systems for wastewater treatment. *Ecol Eng* 25:478–490
27. Vymazal J, Greenway M, Tonderski K, Brix H, Mander Ü (2006) Constructed wetlands for wastewater treatment. In: Verhoeven JTA, Beltman B, Bobbink R, Whigham DF (eds) *Wetlands and natural resource management*, vol 190, *Ecological Studies*. Springer, Berlin/Heidelberg, pp 69–94
28. Wallace SD, Knight RL (2006) Small scale constructed wetland treatment systems. Feasibility, design criteria, and O&M requirements. Water Environment Research Foundation, Alexandria
29. Kadlec RH, Knight RL, Vymazal J, Brix H, Cooper PF, Haberl R (2000) *Constructed wetlands for water pollution control: processes, performance, design and operation*. International Water Association Scientific and Technical report no. 8, London
30. Rogers HH, Davis DE (1972) Nutrient removal by waqter hyacinth. *Weed Sci* 20:423–427
31. Penfound WT, Earle TT (1948) The biology of the water hyacinth. *Ecol Monogr* 18:447–472
32. Crites R, Tchobanoglous G (1998) *Small and decentralized wastewater management systems*. McGraw-Hill, Boston
33. Culley DD Jr, Epps EA (1973) Use of duckweed for waste treatment and animal feed. *J Water Pollut Control Fed* 45:337–347
34. Rai DN, Munshi JD (1979) The influence of thick floating vegetation (waterhyacinth-*Eichhornia crassipes*) on the physico-chemical environment of a freshwater wetland. *Hydrobiologia* 62:65–69
35. Wolverton BC (1979) Engineering design data for small vascular aquatic plant wastewater treatment systems. In: Bastian RK, Reed SC (eds) *Aquaculture systems for wastewater treatment: seminar proceedings and engineering assessment*. EPA 430/9-80-006. U.S. EPA, Washington, pp 179–192
36. Reddy KR, DeBusk WF (1984) Growth characteristics of aquatic macrophytes cultured in nutrient-enriched water. I. Water hyacinth, water lettuce and pennywort. *Econ Bot* 38:229–239
37. Rejmánková E, Rejmánek M, Květ J (1990) Minimizing duckweed (Lemnaceae) production by suitable harvest strategy. In: Whigham DF, Good RE, Květ J (eds) *Wetland ecology and management: case studies*. Kluwer, Dordrecht, pp 39–43
38. Reed SC, Middlebrooks EJ, Crites RW (1995) *Natural systems for waste management and treatment*, 2nd edn. McGraw-Hill, New York

39. Vymazal J (2008) Constructed wetlands, surface flow. In: Jørgensen SE (ed) Encyclopedia of ecology, vol 1. Elsevier B.V, Amsterdam, pp 765–777
40. Stewart EA III, Haselow DL, Wyse NM (1987) Review of operations and performance data of five water hyacinth based treatment systems in Florida. In: Reddy KR, Smith WH (eds) Aquatic plants for water treatment and resource recovery. Magnolia, Orlando, pp 279–288
41. Ozimek T, Czyprzński P (2003) Ten years' experience of constructed wetlands in Poland. Publ Inst Geogr Univ Tartuensis 94:163–167
42. Denny P (1980) Solute movement in submerged angiosperms. Biol Rev 55:65–92
43. Carignan R, Kalff J (1980) Phosphorus sources for aquatic weeds: water or sediments? Science 207:987–989
44. Twilley RR, Kemp WM, Staver KW, Stevenson JC, Boynton WR (1985) Nutrient enrichment of estuarine submersed vascular plant communities. 1. Algal growth and effects on production of plants and associated communities. Mar Ecol Prog Ser 23:179–191
45. Gu B, DeBusk TA, Dierberg FE, Chimney MJ, Pietro KC, Aziz T (2001) Phosphorus removal from Everglades agricultural runoff by submerged aquatic vegetation/limerock treatment technology: an overview of research. Water Sci Technol 44(11/12):101–108
46. Bishop PL, Eighmy TT (1989) Aquatic wastewater treatment using *Eloдея nuttallii*. J Water Pollut Control Fed 61: 641–648
47. Toet S, Van Logtestijn RSP, Schreier M, Kampf R, Verhoeven JTA (2005) The functioning of a wetland system used for polishing effluent from a sewage treatment plant. Ecol Eng 25:101–124
48. Yang Y, Zhencheng X, Kangping H, Junsan W, Guizhi W (1994) Removal efficiency of the constructed wetland wastewater treatment system at Bainikeng, Shenzhen. In: Proceedings of the 4th international conference wetland systems for water pollution control. ICWS'94 Secretariat, Guangzhou, P.R. China, pp 94–103
49. Kadlec RH (1994) Overview: surface flow constructed wetlands. In: Proceedings of the 4th international conference wetland systems for water pollution control. ICWS Secretariat, Guangzhou, P.R.China, pp 1–12
50. Queensland Department of Natural Resources (QDNR) (2000) Guidelines for using freshwater surface constructed wetlands to treat Municipal Sewage. QDNR, Brisbane
51. Vymazal J, Brix H, Cooper PF, Haberl R, Perfler R, Laber J (1998) Removal mechanisms and types of constructed wetlands. In: Vymazal J, Brix H, Cooper PF, Haberl R, Perfler R (eds) Constructed wetlands for wastewater treatment in Europe. Backhuys, Leiden, pp 17–66
52. Richardson CJ (1985) Mechanisms controlling phosphorus retention capacity in freshwater wetlands. Science 228:1424–1427
53. Sundblad K, Wittgren H-B (1989) *Glyceria maxima* for wastewater nutrient removal and forage production. Biol Waste 27:29–42
54. Gersberg RM, Gearhart RA, Ives M (1989) Pathogen removal in constructed wetlands. In: Hammer DA (ed) Constructed wetlands for wastewater treatment. Lewis, Chelsea, pp 431–446
55. Lakatos G (1998) Hungary. In: Vymazal J, Brix H, Cooper PF, Green MB, Haberl R (eds) Constructed wetlands for wastewater treatment in Europe. Backhuys, Leiden, pp 191–206
56. Ewel KC, Odum HT (eds) (1984) Cypress swamps. University of Florida Press, Gainesville
57. Kadlec RH, Tilton DL (1979) The use of freshwater wetlands as a tertiary wastewater treatment alternative. CRC Crit Rev Environ Control 9:185–212
58. Litchfield DK, Schatz DD (1989) Constructed wetlands for wastewater treatment at Amoco Oil Company's Mandan, North Dakota refinery. In: Hammer DA (ed) Constructed wetlands for wastewater treatment. Lewis, Chelsea, pp 233–237
59. Chan E, Bursztynsky TA, Hantzsch N, Litwin YJ (1982) The use of wetlands for water pollution control. U.S. EPA Report EPA-600/2-82-086, Municipal Environmental Research Lab, Cincinnati
60. Wieder RK (1989) A survey of constructed wetlands for acid coal mine drainage treatment in the eastern United States. Wetlands 9:299–315
61. Kadlec RH (2003) Status of treatment wetlands in North America. In: Dias V, Vymazal J (eds) Proc. Conf. The use of aquatic macrophytes for wastewater treatment in constructed wetlands. ICN and INAG, Lisbon, pp 363–401
62. Greenway M, Woolley A (1999) Constructed wetlands in Queensland: performance efficiency and nutrient bioaccumulation. Ecol Eng 12:39–55
63. Tanner CC, Nguyen ML, Sukias JPS (2005) Nutrient removal by a constructed wetland treating subsurface drainage from a grazed dairy pasture. Agric Ecosyst Environ 105:145–162
64. Kyambadde J, Kansime F, Dalhammar G (2005) Nitrogen and phosphorus removal in substrate-free pilot constructed wetlands with horizontal surface flow in Uganda. Water Air Soil Pollut 165:37–59
65. Van Oostrom AJ (1995) Nitrogen removal in constructed wetlands treating nitrified meat processing effluent. Water Sci Technol 32(3):137–148
66. Brix H (1987) Treatment of wastewater in the rhizosphere of wetland plants – the root zone method. Water Sci Technol 19:107–118
67. Vymazal J (2001) Types of constructed wetlands for wastewater treatment: their potential for nutrient removal. In: Vymazal J (ed) Transformations of nutrients in natural and constructed wetlands. Backhuys, Leiden, pp 1–93
68. Kickuth R (1982) A low-cost process for purification of municipal and industrial waste water. Der Tropenlandwirt 83: 141–154
69. Kickuth R (1984) Das Wurzelraumverfahren in der praxis. Landschaft und Stadt 16:145–153
70. Haberl R, Perfler R (1990) Seven years of research work and experience with wastewater treatment by a reed bed system. In: Cooper PF, Findlater BC (eds) Constructed wetlands in water pollution control. Pergamon Press, Oxford, pp 205–214

71. Cooper PF, Green MB (1995) Reed bed treatment systems for sewage treatment – the first 10 years experience. *Water Sci Technol* 32:317–327
72. Vymazal J, Brix H, Cooper PF, Green MB, Haberl R (eds) (1998) *Constructed wetlands for wastewater treatment in Europe*. Backhuys, Leiden
73. Fetter CW, Sloey WE, Spangler FL (1976) Potential replacement of septic tank drain fields by artificial marsh wastewater treatment systems. *Ground Water* 14:1–7
74. Spangler FL, Sloey WE, Fetter CW Jr (1976) Artificial and natural marshes as wastewater treatment systems in Wisconsin. In: Tilton DL, Kadlec RH, Richardson CJ (eds) *Freshwater wetlands and sewage effluent disposal*. University of Michigan, Ann Arbor, pp 215–240
75. Small M, Wurm C (1977) Data report. Meadow/Marsh/Pond System. Brookhaven National Laboratory, BNL 50675
76. Wolverton BC (1987) Artificial marshes for wastewater treatment. In: Reddy KR, Smith WH (eds) *Aquatic plants for water treatment and resource recovery*. Magnolia, Orlando, pp 141–152
77. Gersberg RM, Elkins BV, Goldman CR (1983) Nitrogen removal in artificial wetlands. *Water Res* 17:1009–1014
78. Steiner GR, Watson JT, Hammer DA, Harker DF Jr (1987) Municipal wastewater treatment with artificial wetlands: a TVA/Kentucky demonstration. In: Reddy KR, Smith WH (eds) *Aquatic plants for water treatment and resource recovery*. Magnolia, Orlando, pp 923–932
79. Finlayson CM, Chick AJ (1983) Testing the potential of aquatic plants to treat abattoir effluent. *Water Res* 17:415–422
80. Bavor HJ, Roser DJ, McKersie S (1987) Nutrient removal using shallow lagoon-solid matrix macrophyte systems. In: Reddy KR, Smith WH (eds) *Aquatic plants for water treatment and resource recovery*. Magnolia, Orlando, pp 227–235
81. Davison L, Pont D, Bolton K, Headley T (2006) Dealing with nitrogen in subtropical Australia: seven case studies in the diffusion of ecotechnological innovation. *Ecol Eng* 28:213–223
82. Wood A (1990) The application of artificial wetlands in South Africa. In: Cooper PF, Findlater BC (eds) *Constructed wetlands in water pollution control*. Pergamon Press, Oxford, pp 235–244
83. Byekwaso E, Kansime F, Logstrup J, Andersen S (2002) The optimisation of a reed bed filter for effluent treatment at Kasese Cobalt Company Limited – Uganda. In: *Proceedings of the 8th international conference wetland systems for water pollution control*. University of Dar es Salaam, Tanzania, pp 660–668
84. Kaseva ME (2004) Performance of a sub-surface flow constructed wetland in polishing pre-treated wastewater – a tropical case study. *Water Res* 38:681–687
85. Wang J, Cai X, Chen Y, Yang Y, Liang M, Zhang Y, Wang Z, Li Q, Liao X (1994) Analysis of the configuration and the treatment effect of constructed wetland wastewater treatment system for different wastewaters in South China. In: *Proceedings of the 4th international conference wetland systems for water pollution control*. ICWS'94 Secretariat, Guangzhou, P.R. China, pp 114–120
86. Juwarkar AS, Verma M, Meshram J, Bal AS, Juwarkar A (1992). Wastewater treatment in constructed wetlands. In: *Proceedings of the 3rd international conference wetland systems for water pollution control*. IWA and University of Western Sydney, pp 35.1–35.4
87. Lee CY, Lee CC, Lee FY, Tseng SK, Liao CJ (2004) Performance of subsurface flow constructed wetland taking pretreated swine effluent under heavy loads. *Bioresour Technol* 92:173–179
88. Lin YF, Jing SR, Lee DY, Chang YF, Chen YM, Shih KC (2005) Performance of a constructed wetland treating intensive shrimp aquaculture wastewater under high hydraulic loading rate. *Environ Pollut* 134:411–421
89. Dallas S, Ho G (2005) Subsurface flow reedbeds using alternative media for the treatment of domestic greywater in Monteverde, Costa Rica, Central America. *Water Sci Technol* 51(10):119–128
90. Platzer M, Cáceres V, Fong N, Haberl R (2002) Investigations and experiences with subsurface flow constructed wetlands in Nicaragua, Central America. In: *Proceedings of the 8th international conference on wetland systems for water pollution control*. University of Dar es Salaam and IWA, Arusha, Tanzania, pp 350–365
91. Cooper PF, Boon AG (1987) The use of *Phragmites* for wastewater treatment by the Root Zone Method: the UK approach. In: Reddy KR, Smith WH (eds) *Aquatic plants for water treatment and resource recovery*. Magnolia, Orlando, pp 153–174
92. Young TC, Collins AG, Theis TL (2000) Subsurface flow wetland for wastewater treatment at Minoa. Report to NYSERDA and U.S. EPA, Clarkson University, New York
93. Brix H (1997) Do macrophytes play a role in constructed treatment wetlands? *Water Sci Technol* 35(5):11–17
94. Drew MC (1979) Plant responses to anaerobic conditions in soil and solution culture. *Curr Adv Plant Sci* 11: 36.1–36.14
95. Vymazal J (2007) Removal of nutrients in various types of constructed wetlands. *Sci Total Environ* 380:48–65
96. Vohla C, Pöldvere E, Noorvee A, Kuusemets V, Mander Ü (2005) Alternative filter media for phosphorus removal in a horizontal subsurface flow constructed wetland. *J Environ Sci Health* 40:1251–1264
97. Vymazal J (2005) Removal of enteric bacteria in constructed treatment wetlands with emergent macrophytes: a review. *J Environ Sci Health* 40A:1355–1367
98. Decamp O, Warren A (2000) Investigation of *Escherichia coli* removal in various designs of subsurface flow wetlands used for wastewater treatment. *Ecol Eng* 14:293–299
99. Greiner RW, de Jong J (1984) The use of Marsh Plants for the treatment of waste water in areas designated for recreation and tourism. RIJP report no. 225, Lelystad, The Netherlands
100. Cooper PF (2005) The performance of vertical flow constructed wetland systems with special reference to the significance of oxygen transfer and hydraulic loading rates. *Water Sci Technol* 51(9):81–90

101. Weedon CN (2003) Compact vertical flow reed bed system – first two years performance. *Water Sci Technol* 48(5):15–23
102. Molle P, Liénard A, Boutin C, Merlin G, Iwema A (2005) How to treat raw with constructed wetlands: an overview of French systems. *Water Sci Technol* 51(9):11–21
103. Salati E Jr, Salati E, Salati E (1999) Wetland projects developed in Brazil. *Water Sci Technol* 40(3):19–25
104. Boutin C (1987) Domestic wastewater treatment in tanks planted with rooted macrophytes: case study, description of the system, design criteria, and efficiency. *Water Sci Technol* 19(10):29–40
105. Burka U, Lawrence P (1990) A new community approach to wastewater treatment with higher water plants. In: Cooper PF, Findlater BC (eds) *Constructed wetlands in water pollution control*. Pergamon Press, Oxford, pp 359–371
106. Mæhlum T, Stålnacke P (1999) Removal efficiency of three cold-climate constructed wetlands treating domestic wastewater: effects of temperature, seasons, loading rates and input concentrations. *Water Sci Technol* 40(3):273–281
107. Rousseau DPL, Vanrolleghem PA, De Pauw N (2004) Constructed wetlands in Flanders: a performance analysis. *Ecol Eng* 23:151–163
108. Brix H, Johansen NH (1999) Treatment of domestic sewage in a two-stage constructed wetland - design principles. In: Vymazal J (ed) *Nutrient cycling and retention in natural and constructed wetlands*. Backhuys, Leiden, pp 155–163
109. Brix H, Arias CA, Johansen NH (2003) Experiments in a two-stage constructed wetland system: nitrification capacity and effects of recycling on nitrogen removal. In: Vymazal J (ed) *Wetlands-nutrients, metals and mass cycling*. Backhuys, Leiden, pp 237–258
110. Masi F, Conte G, Martinuzzi N, Pucci B (2002) Winery high organic content wastewaters treated by constructed wetlands in Mediterranean climate. In: *Proceedings of the 8th international conference wetland systems for water pollution control*. University of Dar-es-Salaam, Tanzania and IWA, pp 274–282
111. Mander Ü, Teiter S, Löhmus K, Mäuring T, Nurk K, Augustin J (2003) Emission rates of N₂O and CH₄ in riparian alder forest and subsurface flow constructed wetland. In: Vymazal J (ed) *Wetlands: nutrients, metals and mass cycling*. Backhuys, Leiden, pp 259–279
112. Brix H, Koottatep T, Laugesen CH (2006) Re-establishment of wastewater treatment in tsunami affected areas of Thailand by the use of constructed wetlands. In: Dias V, Vymazal J (eds) *Proceedings of the 10th international conference wetland systems for water pollution control, MAOTDR 2006*, Lisbon, Portugal, pp 59–67
113. Vymazal J (2009) The use constructed wetlands with horizontal sub-surface flow for various types of wastewater. *Ecol Eng* 35:1–17
114. Vymazal J, Kröpfelová L (2008) *Wastewater treatment in constructed wetlands with horizontal sub-surface flow*. Springer, Dordrecht
115. Huang Y, Latorre A, Barcelo D, Garcia J, Aguirre P, Muñeriego R, Bayona JM (2004) Factors affecting linear alkylbenzene sulfonates removal in subsurface flow constructed wetlands. *Environ Sci Technol* 38:2657–2663
116. Matamoros V, Arias C, Brix H, Bayona JM (2009) Preliminary screening of small-scale domestic wastewater treatment systems for removal of pharmaceutical and personal care products. *Water Res* 43:55–62
117. Vymazal J, Kröpfelová L (2009) Removal of organics in constructed wetlands with horizontal sub-surface flow: a review of the field experience. *Sci Total Environ* 407:3911–3922
118. Shutes RBE, Revitt DM, Scholes LNL, Forshaw M, Winter B (2001) An experimental constructed wetland system for the treatment of highway runoff in the UK. *Water Sci Technol* 44(11–12):571–578
119. Headley TR, Huett DO, Davison L (2001) The removal of nutrients from plant nursery irrigation runoff in sub-surface horizontal-flow wetlands. *Water Sci Technol* 44(11–12):77–84
120. Worrall P, Revitt DM, Prickett G, Brewer D (2002) Constructed wetlands for airport runoff – the London Heathrow experience. In: Nehring KW, Brauning SE (eds) *Wetlands and remediation II*. Battelle Press, Columbus, pp 177–186
121. Urbanc-Berčič O (1997) Constructed wetlands for the treatment of landfill leachates: the Slovenian Experience. *Wetlands Ecol Manage* 4:189–197
122. Mæhlum T, Warner WS, Stålnacke P, Jenssen PD (1999) Leachate treatment in extended aeration lagoons and constructed wetlands in Norway. In: Mulamoottil G, McBean EA, Revers F (eds) *Constructed wetlands for the treatment of landfill leachates*. Lewis/CRC Press, Boca Raton, pp 151–163
123. Korkusuz EA, Beglioğlu M, Demirel GN (2005) Comparison of the treatment performances of blast furnace slag-based and gravel-based vertical flow wetlands operated identically for domestic wastewater treatment in turkey. *Ecol Eng* 24:187–200
124. Vohla C, Pöldvere E, Noorvee A, Kuusemets V, Mander Ü (2005) Alternative filter media for phosphorous removal in a horizontal subsurface flow constructed wetland. *J Environ Sci Health A* 40:1251–1264
125. Maddison M, Soosaar K, Mäuring T, Mander Ü (2009) The biomass and nutrient and heavy metal content of cattails and reeds in wastewater treatment wetlands for the production of construction material in Estonia. *Desalination* 246:120–128
126. Zurita F, De Anda J, Belmont M (2009) Treatment of domestic wastewater and production of commercial flowers in vertical and horizontal subsurface-flow constructed wetlands. *Ecol Eng* 35:861–869

Books and Reviews

- Cooper PF (ed) (1990) European design and operation guidelines for reed bed treatment systems. Prepared for the European Community/European Water Pollution Control Association Emergent Hydrophyte Treatment System Expert Contact Group. WRc report UI 17
- Cooper PF, Findlater BC (eds) (1990) Constructed wetlands in water pollution control. Pergamon Press, Oxford
- Cooper PF, Job GD, Green MB, Shutes RBE (1996) Reed beds and constructed wetlands for wastewater treatment. Water Research Center, Swindon
- Hammer DA (ed) (1989) Constructed wetlands for wastewater treatment. Lewis, Chelsea
- Kadlec RH, Knight RL (1996) Treatment wetlands. CRC Press, Boca Raton
- Kadlec RH, Wallace SD (2008) Treatment wetlands, 2nd edn. CRC Press, Boca Raton
- Mander Ü, Jenssen P (eds) (2003) Constructed wetlands for wastewater treatment in cold climates. WIT Press, Southampton
- Mitsch WJ, Gosselink JG (2007) Wetlands, 4th edn. Wiley, Hoboken
- Mitsch WJ, Gosselink JG, Anderson CJ, Zhang L (2009) Wetland ecosystems. Wiley, Hoboken
- Moshiri GA (ed) (1993) Constructed wetlands for water quality improvement. CRC Press/Lewis, Boca Raton
- Mulamoottil G, McBean AE, Revers F (eds) (1999) Constructed wetlands for the treatment of landfill leachates. Lewis/CRC Press, Boca Raton
- Price T, Probert D (1997) Role of constructed wetlands in environmentally-sustainable developments. Appl Energy 57:129–174
- Reddy KR, Smith WH (eds) (1987) Aquatic plants for wastewater treatment and resource recovery, Magnolia Publishing. Orlando, Florida
- Scholz M (2006) Wetland systems to control urban runoff. Elsevier, Amsterdam
- Tinner RW (1999) Wetland indicators. A guide to wetland identification, delineation, classification, and mapping. Lewis, Boca Raton
- Tourbier J, Pierson RW (eds) (1976) Biological control of water pollution. University of Pennsylvania Press, Philadelphia
- Vymazal J (ed) (1999) Nutrient cycling and retention in natural and constructed wetlands. Backhuys, Leiden
- Vymazal J (ed) (2001) Transformations of nutrients in natural and constructed wetlands. Backhuys, Leiden
- Vymazal J (ed) (2003) Wetlands – nutrients, metals and mass cycling. Backhuys, Leiden
- Vymazal J (ed) (2005) Natural and constructed wetlands: nutrients, metals and management. Backhuys, Leiden
- Vymazal J (2008) Wastewater treatment, plant dynamics and management in constructed and natural wetlands. Springer, Dordrecht
- Vymazal J, Kröpfelová L (2008) Wastewater treatment in constructed wetlands with horizontal sub-surface flow. Springer, Dordrecht

Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management

EUNKYUNG LEE¹, SUNGYUN LEE¹, YUNAH KIM¹,
YU-JEONG HUH², KAP-SOON KIM², JAEWEON CHO³

¹Department of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), Buk-gu, Gwangju, South Korea

²Yeongsan River Environment Research Center, Buk-gu, Gwangju, South Korea

³Gwangju Institute Science and Technology (GIST), Department of Environmental Science and Engineering, Natural Organic Matter Lab, Gwangju, Korea

Article Outline

Glossary
Definition of the Subject
Introduction
Materials and Methods
Results and Discussion
Conclusion
Future Directions
Acknowledgments
Bibliography

Glossary

Micropollutant Pollutant that exists in very small traces in aquatic environment including pharmaceutical drugs and endocrine-disturbing compounds.

Pharmaceuticals Chemical substances used in the medical diagnosis, cure, treatment, or prevention of disease.

Metabolites Intermediates and products formed by the metabolic transformations of a drug.

Personal care products Products used by individuals for personal cosmetic reasons, including fragrances and cosmetics.

Anthropogenic indicator Molecular markers of sewage-derived contaminants which is useful for

The two authors (Eunkyoung Lee, Sungyun Lee) are equally contributing first authors.

identifying the sources of pollutants and tracing their transport pathways in aquatic environment.

Wastewater effluent Wastewater discharge from the wastewater treatment systems based upon physical, chemical, and biological processes.

Octanol-water partitioning coefficient Distribution equilibrium of a compound between water and *n*-octanol (K_{ow}) which is helpful to predict adsorption tendency of a compound.

Acidity constant Dissociation equilibrium constant for acid–base reactions, that is, acid dissociation constant.

Definition of the Subject

Micropollutants including pharmaceuticals, personal care products (PPCPs) have been an emerging issue in aquatic environment for several decades. Based on the improvement of the technology to detect and quantify these chemicals, the occurrence, fate, and removal mechanism of various micropollutants has extensively been investigated through different physical, chemical, and biological water treatments in different water sources. Rivers, lakes, estuarine waters, groundwaters, and even drinking waters are subject to the contaminations resulting from the release of anthropogenic micropollutants from wastewater treatment plants. The ubiquitous distribution of PPCPs and their metabolites is mainly caused by the discharge of partially eliminated wastewater contaminants into rivers, with growing concerns of potential ecotoxicological effects of pharmaceuticals on environment. With this point, the micropollutants have to be monitored for sustainable water management. Besides the existence, micropollutants with varying chemical characteristics, in terms of acidity constant (pK_a), octanol-water partitioning coefficient ($\log K_{ow}$), may exhibit different transport behaviors in aquatic environment, with those being divided into conservative versus labile markers. There is a great need to suggest anthropogenic micropollutant indicators to predict the levels and behaviors of other micropollutants even with different physicochemical properties, for the sustainable management of the river, as measurements of all the micropollutants require complicated procedures and extensive time and expense.

Introduction

Micropollutants including pharmaceuticals and personal care products (PPCPs) have been extensively

examined by a large number of research groups [1–6]. The worldwide occurrence of PPCPs has not been limited to wastewater plant effluents [2, 7], but have also been observed in river water [5, 8–10], groundwater [7], estuarine water [4, 12], and even drinking water [2, 13]. In an attempt to reduce the anthropogenic burden of these pollutants on the environment, recent studies have focused on the removal of micropollutants, including pharmaceuticals, by means of activated sludge treatment, bank filtration, oxidation, and wetland treatment [14, 15]. Some previous studies have also suggested that the existence of certain micropollutants could be used as anthropogenic molecular markers to trace wastewater-derived contaminants [4, 7, 16, 17].

According to Nakada et al. [4], in riverine catchments in Japan, there were relatively high correlations between the levels of crotamiton and carbamazepine and human population compared to biodegradable (ibuprofen), photodegradable (naproxen), volatile, and hydrophobic (triclosan) compounds. Some studies have attempted to use pharmaceuticals, especially those being used solely by humans, such as carbamazepine, diphenhydramine, and caffeine, as chemical indicators of human fecal contamination in water samples upstream and downstream of wastewater treatment plants and the corresponding wastewater effluent [7]. Clara et al. [18] also suggested that carbamazepine would be a good anthropogenic marker since it was highly persistent in 11 full-scale wastewater treatment plants in Austria, regardless of the wastewater treatment processes, including activated sludge, trickling filter, nitrification, denitrification, and phosphorus removal systems. The levels of caffeine, nitrate, and ammonia were also investigated in many different groundwater wells that were connected to either wastewater plants or septic tanks to test the potential of using caffeine as a wastewater discharge indicator. The results of this study demonstrated both the possibility and limitation of using caffeine as the indicator since the amount of discharged caffeine was very high and caffeine was not susceptible to biodegradation [19].

Several recent studies have been conducted to examine the potential fate and transport mechanisms of pharmaceuticals and to correlate these factors to wastewater contaminants [10, 20]. Correlation analysis with the representative parameters of wastewater

contaminants has provided statistically approved indicators. Yoon et al. [10] reported that the COD_{cr} was strongly correlated with EDC/PPCP concentrations ($r = 0.78$), and Santos et al. [20] reported that pharmaceuticals were positively correlated to wastewater characterization parameters such as TSS, BOD, and COD.

The hydrophobicity parameter, $\log K_{\text{ow}}$, is believed to be an important indicator for specific compounds in predicting their bio-resistance and adsorption tendency, and consequently this parameter can be utilized to evaluate pharmaceuticals as potential markers of contamination in riverine and estuarine ecosystems. In a previous study conducted by Glassmeyer et al. [7], the $\log K_{\text{ow}}$ parameters of selected pharmaceuticals, including carbamazepine, diphenhydramine, and caffeine, were not examined, but it is believed that information on the $\log K_{\text{ow}}$ of the tested compounds can help explain their behaviors in aquatic ecosystems. The $\log K_{\text{ow}}$ value (-0.07) of caffeine is relatively low, thus it is highly biodegradable; meanwhile, the values of carbamazepine and diphenhydramine are relatively high (2.45 and 3.20) and the $\log K_{\text{ow}}$ value of diphenhydramine is dependent on pH. These factors suggest that carbamazepine holds great promise as an effective human fecal indicator, which is similar to the results reported in other studies. Functional groups, such as the carboxylic moiety, may substantially change the $\log K_{\text{ow}}$ to normally lower values at neutral pH (e.g., 3.79–0.80 at pH 7.4 for ibuprofen with carboxylic group). The biodegradation potential of pharmaceuticals in constructed-engineered wetlands fed with wastewater effluent has been shown to be tightly related to the pH when considered in conjunction with $\log K_{\text{ow}}$ (i.e., $\log D$) [21]. The $\log D$ parameter has rarely been applied in micropollutant research, but is believed to provide important information in predicting the fate of micropollutants in an aquatic environment.

The objectives of this study were: First, to investigate the occurrence of selected micropollutants including pharmaceuticals and their metabolites in Yeongsan River water; second, to evaluate potential anthropogenic molecular markers in the river water by correlating their levels to wastewater quality parameters; and lastly, to elucidate the behaviors of pharmaceuticals with regard to the chemical characteristics, $\log D$ and pK_{a} .

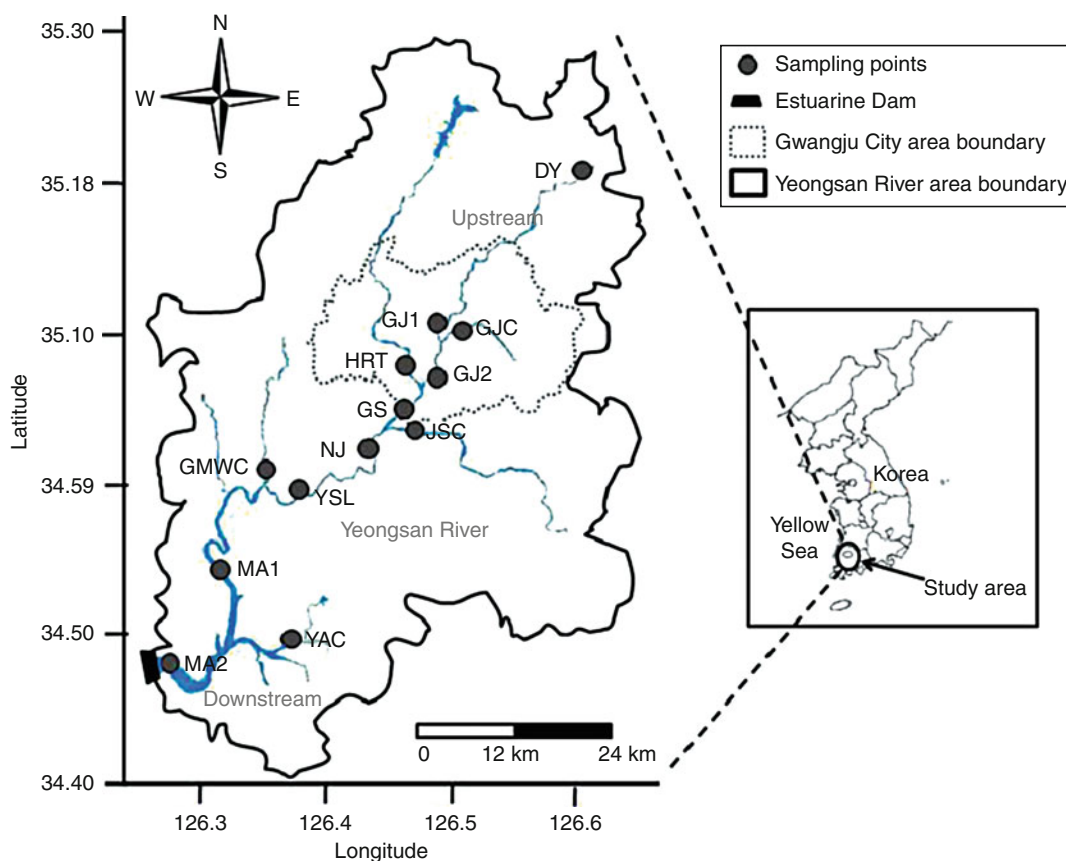
Materials and Methods

Investigated Area and Sampling Sites

The Yeongsan River located in southwestern Korea was selected to investigate the possibility of using certain pharmaceuticals and personal care products (PPCPs) as anthropogenic indicators. The Yeongsan River system covers a basin area of 3,371 km² and has a total length of 115.5 km. The Yeongsan River starts from Damyang County, runs through three big cities, Gwangju, Naju, and Mokpo cities, and comes to the Yellow Sea through the estuarine barrier. Gwangju discharges a significant amount of wastewater to the Yeongsan River area, and has a population of approximately 1.4 million inhabitants. This city also contains several wastewater treatment plants that utilized the activated sludge process and discharges between 120,000 and 600,000 m³/day. The Gwangju city is most likely the most important source of anthropogenic micropollutants, due to its dense population relative to adjacent cities around the Yeongsan River. As shown in Fig. 1, a total of 13 sampling sites upstream and downstream of the Yeongsan River were selected for micropollutants analysis. Sample collection was conducted four times from March to October in 2009. The precipitation in Gwangju area was mainly concentrated from June to September implying the dilution of contaminants (World Weather Information Service <http://worldweather.wmo.int/095/c00335.htm>). The weather conditions (water level, precipitation, temperature, and sun irradiation) in Gwangju are shown in Fig. 2. Sampling sites were changed between March and July from a project guideline switching, and maintained for later investigations.

Sample Collection and Water Quality Analysis

1 L river-water samples were collected from the Yeongsan River in amber glass bottles which had been previously cleaned with methanol and deionized pure water. One milliliter of 2 N sodium azide was spiked onsite into the sample bottles to prevent microbial activity. Samples were transported to the laboratory on ice and stored at 4°C until extraction. In situ measurements were conducted in terms of temperature, pH, conductivity, oxidation reduction potential (ORP), and dissolved oxygen (DO). In order to



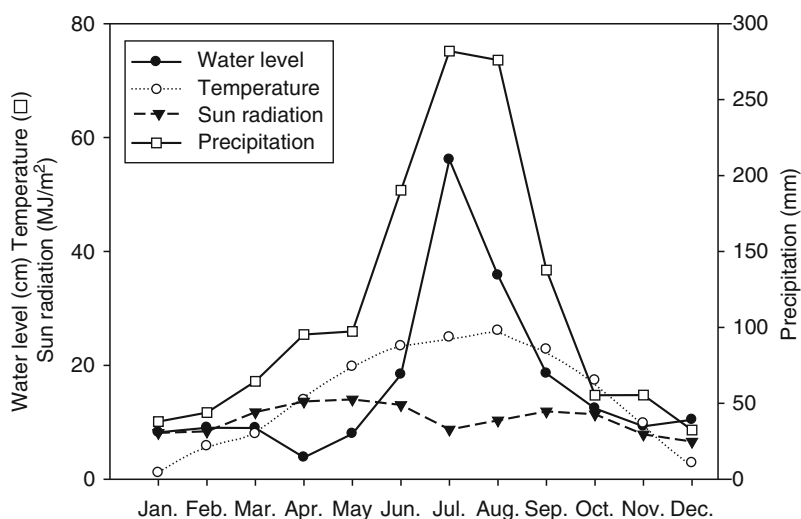
Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Figure 1 Map of sampling sites in Yeongsan River area, Korea (DY Damyang, GJ1 Gwangju 1, GJC Gwangju Creek, GJ2 Gwangju 2, GS Gwangsan, HRT Hwangrong tributary, JSC Jiseok Creek, NJ Naju, YSL Yeongsan Lake, GMWC Gomakweon Creek, MA1 Muan 1, YAC Youngam Creek, MA2 Muan2)

investigate the relationship between water characteristics and micropollutants, the characteristics of the water samples without sodium azide were examined in terms of dissolved organic carbon (DOC), ultraviolet absorbance (UVA), specific UVA (SUVA), and total nitrogen (TN) after pretreatment using 0.7 μm glass macro-fiber filters (GFF, Whatman, Maidstone, England). The level of DOC was measured using a total carbon analyzer (Sievers, Boulder, CO, USA), equipped with an auto sampler, by employing both chemical and UV oxidation methods. DOC for brackish water, that is, non-purgable organic carbon (NPOC), was measured using the TOC V_{CPH} (Shimadzu, Kyoto, Japan), by employing the combustion catalytic oxidation method simultaneously with a nondispersive infrared (NDIR) gas analyzer TN,

which was measured in TN unit (TNM-1, Shimadzu, Japan), equipped with the Shimadzu TOC analyzer. UV absorption at 254 nm was analyzed using a UV-1601 UV spectrophotometer (Shimadzu, Japan). SUVA was obtained by calculating the ratio of UVA to DOC. The fundamental water characteristics of the tested samples are shown in Fig. 3. DOC and TN were used as indicators of discharged contaminants from both point and nonpoint sources. SUVA was used as an indicator of the aromaticity of natural and effluent organic matter.

Target Compounds

Various micropollutant compounds, including 13 pharmaceuticals, four selected metabolites, and one



Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Figure 2 The 2009 monthly averaged temperature, sun radiation, water levels of Yeongsan River, and a 10-year average of precipitations in Gwangju, Korea

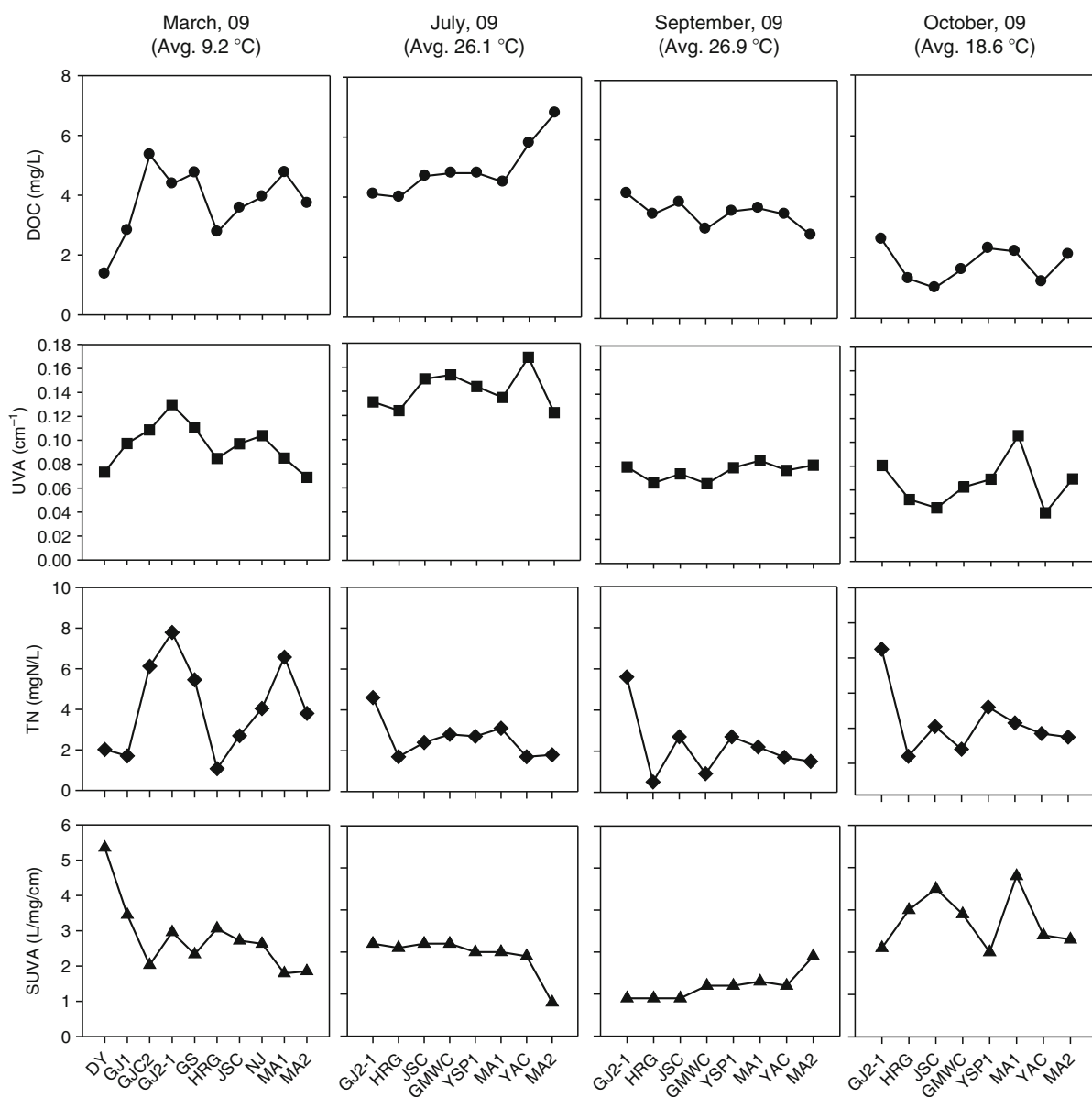
personal care product, were selected as listed in Table 1. Acetaminophen (ACT), atenolol (ATN), carbamazepine (CBZ), clopidogrel (CPG), diclofenac (DCF), dilantin (DLT), florfenicol (FFC), glimepiride (GMP), ibuprofen (IBU), naproxen (NPX), *O*-desmethyl-naproxen (*O*-desmethyl-NPX), sulfamethoxazole (SMX), and tri(2-chloroethyl) phosphate (TCEP) were obtained from Sigma-Aldrich (St. Louis, MO). Caffeine (CAF) was purchased via Fluka Chemie GmbH (Buchs, Switzerland). 2-hydroxyl-ibuprofen (IBU-2OH), ibuprofen carboxylic acid (IBU-CA), iopromide (IOP), and *N*-acetyl-sulfamethoxazole (*N*-acetyl-SMX) were purchased via Tronto Research Inc (Tronto, Canada). Carbamazepine- d_{10} and $^{13}\text{C}_1$ -naproxen- d_3 were purchased from Cambridge Isotope Laboratories (Andover, MA, US). Atenolol- d_7 , diclofenac- d_4 , and dilantin- d_{10} were obtained from C/D/N Isotopes (Pointe-Claire, Canada). Florfenicol- d_3 , ibuprofen- d_3 , N^4 -acetylsulfamethoxazole- d_4 , and sulfamethoxazole- d_4 were obtained from Tronto Research Inc (Tronto, Canada). HPLC grade methanol was obtained from J.T Baker (Philipsburg, NJ, USA). Methyl *tert*-butyl ether (MTBE) and formic acid were obtained from Sigma-Aldrich (St. Louis, MO, USA).

These selected compounds have different aromaticity and hydrophobicity, in terms of both $\log K_{ow}$ and

$\log D$ values at pH 7, and contain ionizable functional groups, which can be measured through pK_a values (Table 1). The relative degree of biodegradability (or persistency) of the tested compounds can be determined by considering both the $\log K_{ow}$ and $\log D$ values. Compounds with $\log K_{ow}$ and $\log D$ values higher than 2.0 (i.e., carbamazepine, TCEP) are highly persistent through biologically active systems, such as wastewater treatment plants and free surface constructed wetlands. Even with relatively high $\log K_{ow}$ values, compounds with lowered $\log D$ (i.e., $\log K_{ow}$ at pH 7) values, have higher removal efficiencies from biologically active ecosystems.

Solid Phase Extraction

The tested micropollutants were extracted using an AutoTrace automated solid phase extraction (SPE) system (Caliper Corporation, Hopkington, MA), equipped with 6 mL, 500 mg hydrophilic-lipophilic balance (HLB) glass cartridges (Waters Corporation, Milford, MA). The procedure used for SPE was described in a previous study by Vanderford et al. [1]. Briefly, the cartridges were preconditioned in the following order: 5 mL of MTBE, 5 mL of methanol, and 5 mL of deionized water. Five hundred milliliters



Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Figure 3 Characteristics of Yeongsan River waters, Korea

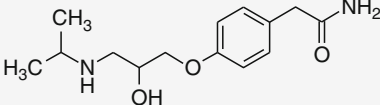
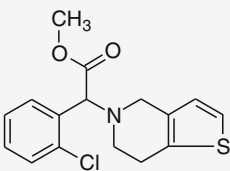
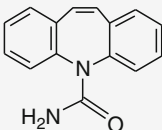
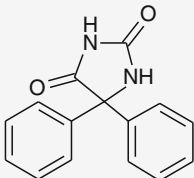
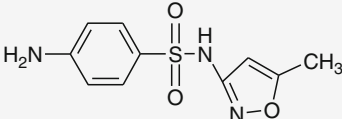
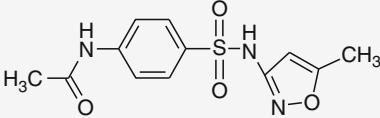
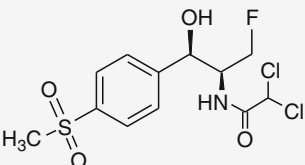
of samples, spiked with the addition of standards for internal calibration, were loaded onto the cartridges at 15 mL/min in duplicate, after which the cartridges were rinsed with 5 mL of deionized pure water, and then dried with a stream of air for 50 min. The cartridges were eluted with 5 mL of methanol, followed by 5 mL of 1/9 (v/v) methanol/MTBE. The eluted solution was concentrated in a water bath at 40°C with a gentle

stream of air to a final volume of 500 μL , which was a concentration factor of 1,000.

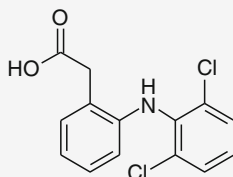
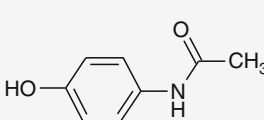
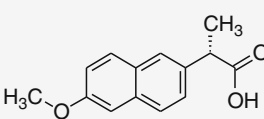
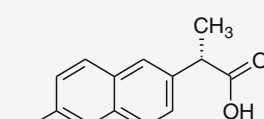
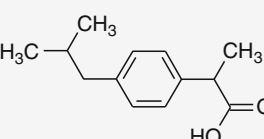
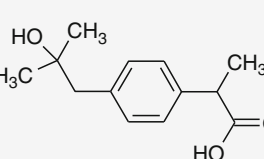
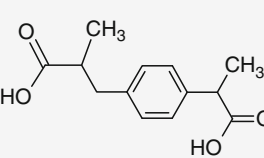
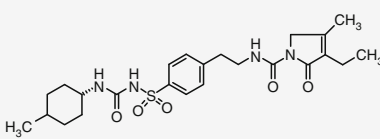
Liquid Chromatography and Electrospray Tandem Mass Spectrometry

Whole analytes were measured using a Water 2,695 Separations Module (Waters, Milford, MA) coupled

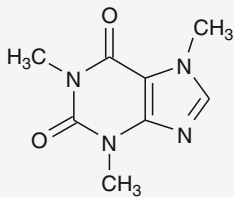
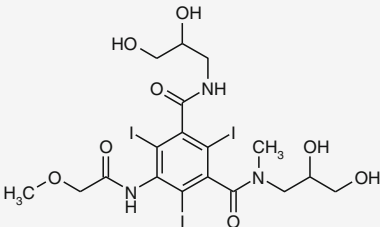
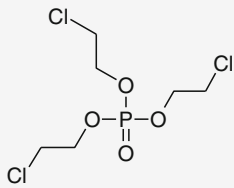
Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Table 1
Selected target pharmaceuticals and personal care product

Analyts	Uses	Structure	pK _a ^a	logK _{ow} ^a	logD ^a at pH7
Atenolol	Beta-blocker		NE ^b	0.43	−2.14
Clopidogrel	Antiplatelet agent		NE	4.03	3.97
Carbamazepine	Anticonvulsant		NE	2.77	2.77
Dilantin	Anticonvulsant		8.46	2.15	2.13
Sulfamethoxazole	Antibiotic		6.16	0.79	0.14
N ⁴ -acetylsulfamethoxazole	Sulfamethoxazole metabolite		5.88	0.86	0.1
Florfenicol	Antibiotic		9.03	0.67	0.67

Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Table 1
(Continued)

Analyts	Uses	Structure	pKa ^a	logK _{ow} ^a	logD ^a at pH7
Diclofenac	Analgesic		4.00	4.26	0.96
Acetaminophen	Analgesic		9.46	0.91	0.91
Naproxen	Analgesic		4.19	2.99	0.25
O-desmethyl-naproxen	Naproxen metabolite		4.34, 9.78	3.9	0.23
Ibuprofen	Analgesic		4.85	3.84	1.71
2-hydroxyl-ibuprofen	Ibuprofen metabolite		4.63	2.08	0.03
Ibuprofen carboxylic acid	Ibuprofen metabolite		3.97, 4.77	2.78	-2.36
Glimepiride	Anticholesterol		4.32	3.12	2.18

Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Table 1 (Continued)

Analyts	Uses	Structure	pK _a ^a	logK _{ow} ^a	logD ^a at pH7
Caffeine	Stimulant		NE	−0.55	−0.55
Iopromide	X-ray contrast media		NE	−1.49	−1.54
tri(2-chloroethyl) phosphate (TCEP)	Flame retardant		NE	2.11	2.11

^alogK_{ow}, log D, pK_a value was calculated from the Software Calculator Plugins^bNE not exist at pH range 1–14

with a Micromass Quattro Micro triple-quadrupole tandem mass spectrometer (Micromass, Manchester, UK) in electrospray ionization mode (ESI). A 20- μ L sample loop and 150 $\times$ 2.1-mm SunFire C18 column with a particle size of 3.5 μ m (Waters, Milford, MA) was employed for analyte separation. A binary gradient consisting of 0.1% formic acid (eluent A) and 100% acetonitrile (eluent B) at a flow rate of 0.2 mL/min was equally applied in both ESI positive and negative modes. For target micropollutants analyses, the gradient was as follows: gradient with 10% of B was held for 0.5 min, stepped up to 50% at 0.51 min, and linearly increased to 100% for 10 min, then held at 100% for 5 min.

The triple-quadrupole tandem mass spectrometer was operated in the multiple reaction monitoring (MRM) mode to identify the analytes. Details on the

MRM data of target compounds are listed in Table 2. The operation conditions were: capillary voltage, 3.0 kV; the source temperature, 150°C; desolvation temperature, 350°C; cone gas, 30 L/h; desolvation gas, 500 L/h; and collision cell pressure, 3 $\times$ 10^{−3} mbar.

Limit of Detection and Calibration

Method detection limit (MDL) was determined using seven different deionized water samples, fortified with unlabeled standards at 5 and 10 ng/L through the same extraction procedure used for the river-water samples. Statistical MDL was calculated by multiplying the appropriate student's T value (degree of freedom is 2.764 for $n - 1 = 10$) by the standard deviation of the replicate measurements [22]. The range of the spiked levels for MDL determination should be included

Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Table 2
Optimized parameters of selected micropollutants

compound	Cone voltage (v)	Collision energy (eV)	Parent ion (m/z)	Daughter ion (m/z)	Retention time (min)
<i>ESI positive</i>					
Atenolol	35	23	267	145	2.25
Atenolol-d ₇	35	25	274	145	2.24
Caffeine	35	20	195.	138	6.11
Sulfamethoxazole	30	18	254	156	7.83
Sulfamethoxazole-d ₄	25	15	258	160	7.80
Dilantin	35	13	253	182	8.68
Dilantin-d ₁₀	30	12	263	192	8.63
Carbamazepine	35	18	237	194	8.70
Carbamazepine-d ₁₀	35	18	247	204	8.74
TCEP	30	16	285	161	9.12
Glimepiride	28	13	491	352	11.34
Clopidogrel	25	20	322	184	12.05
Acetaminophen	28	17	152	109.9	3.98
<i>ESI negative</i>					
Iopromide	20	14	790	127	2.13
Naproxen	10	8	229	185	9.84
Naproxen-d ₃	10	6	233	189	9.83
O-desmethyl-naproxen	15	12	215	171	8.10
Diclofenac	15	10	294	249	11.27
Diclofenac-d ₄	15	12	298	254	11.23
Ibuprofen	15	8	205	161	11.63
Ibuprofen-d ₃	15	8	208	164	11.63
2-hydroxyl-ibuprofen	17	7	221	177	8.17
Ibuprofen carboxylic acid	12	5	235	191	8.04
N-acetyl-sulfamethoxazole	30	13	294	198	7.71
N-acetyl-sulfamethoxazole-d ₄	35	14	298	202	7.65
Florfenicol	11	25	356	336	7.84

between calculated MDL and ten times of calculated MDL. In this entry, the limits of detection (Table 3) were calculated by multiplying the MDL by 3. All analytes were quantified using a 7–8 point calibration

graph at 0.5, 1.0, 5.0, 10, 50, 100, 150, and 200 µg/L, and the concentration of the isotopically labeled compounds was 100 µg/L. Recovery of the labeled internal standards was measured and used for comparison with

Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Table 3 Method detection limits (MDL), reporting limits (RL) for target compound, and ranges in concentrations of PPCPs and their metabolites along Yeongsan River ($n = 33$ for PPCPs, $n = 23$ for metabolites)

	MDL (ng/L)	RL (ng/L)	Min. (ng/L)	25th percentile (ng/L)	Median (ng/L)	75th percentile (ng/L)	Max. (ng/L)	Frequency (%)
Atenolol	1.2	3.4	3.4	5.1	9.5	38.8	217.5	93.9
Clopidogrel	0.6	1.9	2.2	2.8	6.3	8.6	10.4	81.8
Dilantin	1.2	3.7	5.4	8.1	16.4	23.8	31.3	78.8
Carbamazepine	0.7	2.1	2.1	10.6	26.2	77.3	180.4	97.0
Sulfamethoxazole	0.6	1.7	3.2	10.6	24.9	52.1	101.1	100.0
Florfenicol	0.7	2.0	2.5	3.6	8.9	18.7	175.6	30.3
Naproxen	3.3	10.0	11.7	16.1	38.3	84.0	265.9	60.6
Ibuprofen	1.2	3.6	5.0	9.9	25.3	69.3	521.4	66.7
Diclofenac	1.3	3.8	5.9	19.2	35.7	66.5	125.0	39.4
Acetaminophen	0.5	1.5	6.7	15.0	29.4	64.2	148.7	100.0
Glimepiride	1.5	4.6	8.0	10.2	31.5	39.4	52.6	69.7
Caffeine	1.5	4.6	12.5	27.3	74.9	209.3	1506.9	100.0
Iopromide	2.1	6.3	11.9	43.5	109.5	255.1	665.3	93.9
N-acetylsulfamethoxazole	0.9	2.7	3.0	6.5	13.0	16.0	22.3	56.5
O-desmethyl-naproxen	1.2	3.5	4.3	22.1	40.0	131.6	205.0	30.4
2-hydroxyl-ibuprofen	0.9	2.7	5.0	24.2	40.8	50.4	197.4	69.6
Ibuprofen carboxylic acid	2.3	6.8	19.6	13.2	19.8	19.9	19.9	8.7
Tri(2-chloroethyl) phosphate	7.9	23.8	29.5	89.7	136.4	253.9	799.6	97.0

the relative response ratio. The final sample concentrations were obtained after modification using the recoveries of both isotopes and inserted standards.

Statistics: Linear Regression and Correlation Coefficient Analysis

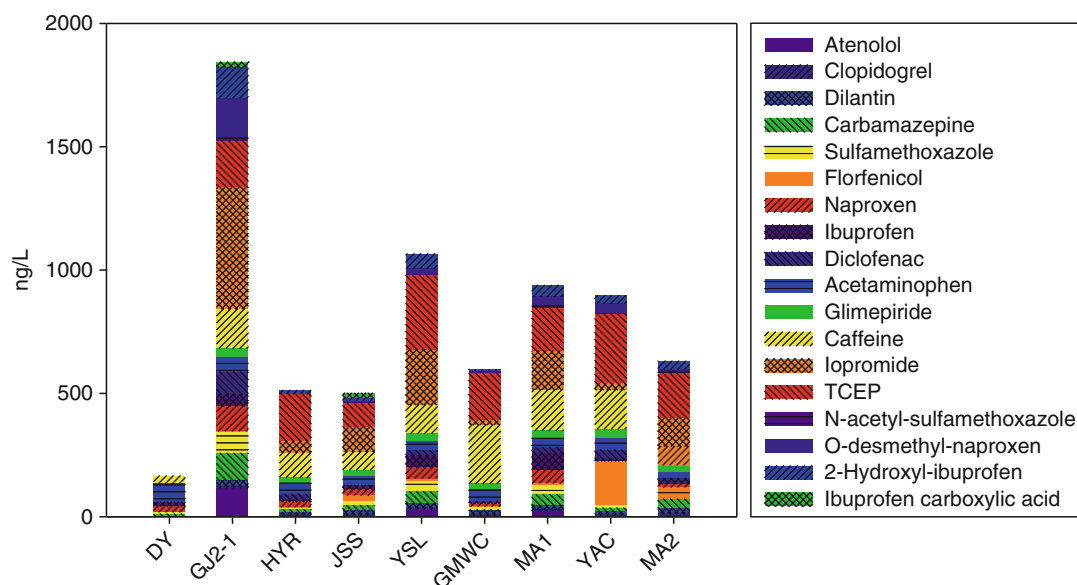
All experimental data were used to create a normal distribution after log-transformation for further statistical analyzes. The data were verified using linear regression analysis between PPCPs and water quality parameters. In addition, correlation analysis between water quality and target compounds was performed by the Pearson correlation. Whole statistical analyses

were conducted using SPSS (Statistics Package for Social Sciences, version 15.0 for windows; SPSS Inc. Chicago, IL, US).

Results and Discussion

Spatial and Temporal Occurrence of Micropollutants During the Sampling Period

The levels of selected PPCPs in the Yeongsan River in Korea were investigated. The spatial distribution of PPCPs and their metabolites in the Yeongsan River are described in Fig. 4 and the concentration ranges of PPCPs and their metabolites are summarized in Table 3. As shown in Fig. 4, significantly higher



Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Figure 4 Spatial distribution of PPCPs and their metabolites in Yeongsan River. Averaged concentrations of target compounds at each sampling point were stacked in vertical bar graph

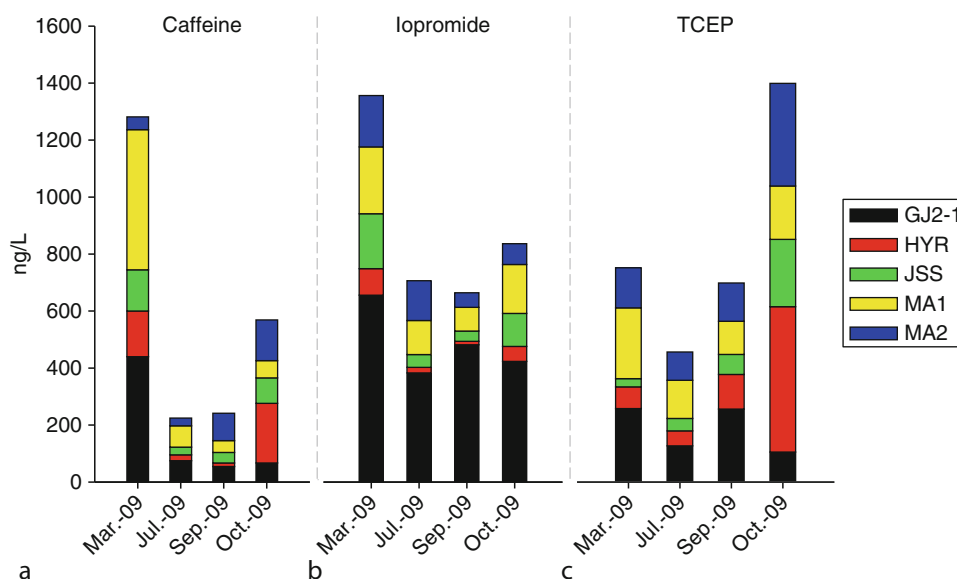
concentrations of most target compounds were observed at sites within the Gwangju city area (GJ2-1) regardless of seasonal variations. On the other hand, at sites upstream and in tributaries of the Yeongsan River (DY, HRR, JSS, and GMWC), most PPCPs were present at relatively low concentrations compared to the concentration in the mainstream of the Yeongsan River, which includes the Gwangju area. For example, the concentration of carbamazepine and iopromide in Gwangju (GJ2-1 site) ranged from 74.2 to 180.4 ng/L and 388.6 to 665.3 ng/L, while the concentrations in samples obtained upstream and in tributaries ranged from 2.1 to 46.1 ng/L and 11.5 to 195.4 ng/L, respectively. Based on these results, Gwangju, which is the major urban center with dense populations, was the major source of anthropogenic micropollutants in the Yeongsan River system.

The occurrence of PPCPs in the river water was found to be affected by temporal variation. Among the various target compounds, caffeine, iopromide, and TCEP exhibited a seasonally dependant behavior throughout all sampling sites on the Yeongsan River (Fig. 5). High concentrations of PPCPs were observed

in March, which drastically decreased in July and September and then increased again in October. These three compounds, which had the highest concentration ranges (median at 70–140 ng/L, Table 3), might be affected by environmental conditions, such as the dilution effect of rainfall [23] in the summer season in Korea (Fig. 2).

Concentrations of Target Compounds in River Water

Most target compounds were successfully detected and their levels were consistent with previous studies [2, 9, 10]. Frequently observed compounds (frequency >90% in Table 3) included atenolol, carbamazepine, sulfamethoxazole, acetaminophen, caffeine, iopromide, and tri(2-chloroethyl) phosphate (TCEP). Among them, only three pharmaceuticals, sulfamethoxazole, caffeine and acetaminophen, were detected at all sampling dates and sites with a frequency of 100% as shown in Table 3. The median (interquartile range) concentration of these three compounds was 24.9 (10.6–52.1) ng/L, 74.9 (27.3–209.3) ng/L, and 29.4 (15.0–64.2) ng/L, respectively. The high level of



Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Figure 5 Temporal distribution of (a) caffeine, (b) iopromide, and (c) TCEP in Yeongsan River

frequency of these compounds was also reported in previous studies [2, 24].

Compounds with relatively high concentrations based on the median values included caffeine, iopromide, and TCEP. The presence of caffeine, which had a median concentration of 74.9 ng/L, has been examined in other surface waters that contain effluents from wastewater treatment plants, such as lakes and rivers, to assess the potential of using it as a chemical anthropogenic marker [17] even though the degradability of caffeine from wastewater treatment plants and surface water systems is high (81.0–99.9% removal in wastewater treatment plants in Swiss).

The X-ray contrast media, iopromide, was also found at significantly higher concentrations than other compounds ranging from 11.9 to 665.3 ng/L (median at 109.5 ng/L). Iopromide has frequently been detected at concentrations of up to 11 µg/L in WWTP effluents and 0.9 µg/L in surface waters due to its persistence and heavy loads [2, 10]. Metabolically stable iopromide, which is known to have a poor inherent biodegradability, exhibited low sorption ability, due to its low logD value of -1.54 at pH 7 (Table 1). Nevertheless, effective biotransformation of iopromide

has been reported using a wastewater treatment system and bank filtration [25].

The flame retardant, TCEP was shown to have the highest median concentration (136.4 ng/L) in the broad concentration ranges of 29.5–799.6 ng/L in the Yeongsan River. Kim et al. [2] reported that TCEP was barely removed during wastewater treatment processes, including membrane bioreactor (MBR), nano-filtration (NF), and even reverse osmosis (RO) filtration system. TCEP was also detected after drinking water treatment processes, which is consistent with their recalcitrant behaviors to chlorine oxidation or ozonation [13]. Therefore, the low removal efficiency of TCEP might be due to its symmetric structure, which may increase its overall stability (Table 1).

Four different anti-inflammatories, including ibuprofen, acetaminophen, diclofenac, and naproxen had concentrations ranging from 20 to 40 ng/L (median) in the Yeongsan River. The concentration range of ibuprofen, acetaminophen, naproxen, and diclofenac was 5.0–521.4 ng/L, 6.7–148.7 ng/L, 11.7–265.9 ng/L, and 5.9–125.0 ng/L, respectively (Table 3). These values were consistent with the combined findings of worldwide studies on surface waters conducted by Fono et al. [8], Herberer [26], Winkler et al. [27], and

Tixier et al. [28]. One of the most popular painkillers in Korea is acetaminophen and this compound has been known to be easily degraded and removed by WWTPs [26]. According to Kolpin et al. [6], acetaminophen was detected at a median concentration of 110 ng/L with a frequency of 23% in US streams. However, in Yeongsan River, acetaminophen was observed at all sampling sites and dates with a frequency of 100%. Diclofenac has been reported to be rapidly and extensively eliminated by photodegradation [29]. Diclofenac was rarely detected in the Yeongsan River, presumably due to photodegradation by sunlight, except in the Gwangju city area, where relatively high concentrations (12.4–100.7 ng/L) of diclofenac were observed due to the dense population and heavy loads from WWTPs. Ibuprofen (median at 25.3 ng/L) and naproxen (median at 38.3 ng/L) were also frequently observed in the Yeongsan River with detection frequencies of 67% and 61%, respectively. Ibuprofen has been known to be relatively persistent in aquatic system [10]. According to Zwiener et al. [30], since ibuprofen, which contains carboxyl groups, had a low removal efficiency by adsorption on activated carbon due to its negative charge in neutral surface water, it is believed to rarely adsorb to suspended particles in river water. However, Tixier et al. [28] reported that ibuprofen can be readily eliminated through biodegradation and sedimentation, and naproxen can be efficiently eliminated through biodegradation and phototransformation in surface waters.

Carbamazepine and sulfamethoxazole were detected in Yeongsan River at concentration ranges of 2.1–180.4 ng/L (median at 26.2 ng/L) and 3.2–101.1 ng/L (median at 24.9 ng/L), respectively. Both compounds have frequently been reported as resistant compounds in surface water, groundwater, and estuarine water [4, 7, 11, 12].

The other compounds that were present at relatively low concentrations were dilantin, atenolol, florfenicol, and clopidogrel, which had median concentrations ranging from 6.3 to 16.4 ng/L (as median). The veterinary antibiotic, florfenicol, has been reported to be rapidly metabolized into florfenicol amine [31], supporting the low concentration of florfenicol detected in this study. Similar to the low concentration range of clopidogrel detected in this study, Huerta-Fontela et al. [32] also found that clopidogrel was rarely detected in WWTP influent and effluent samples in

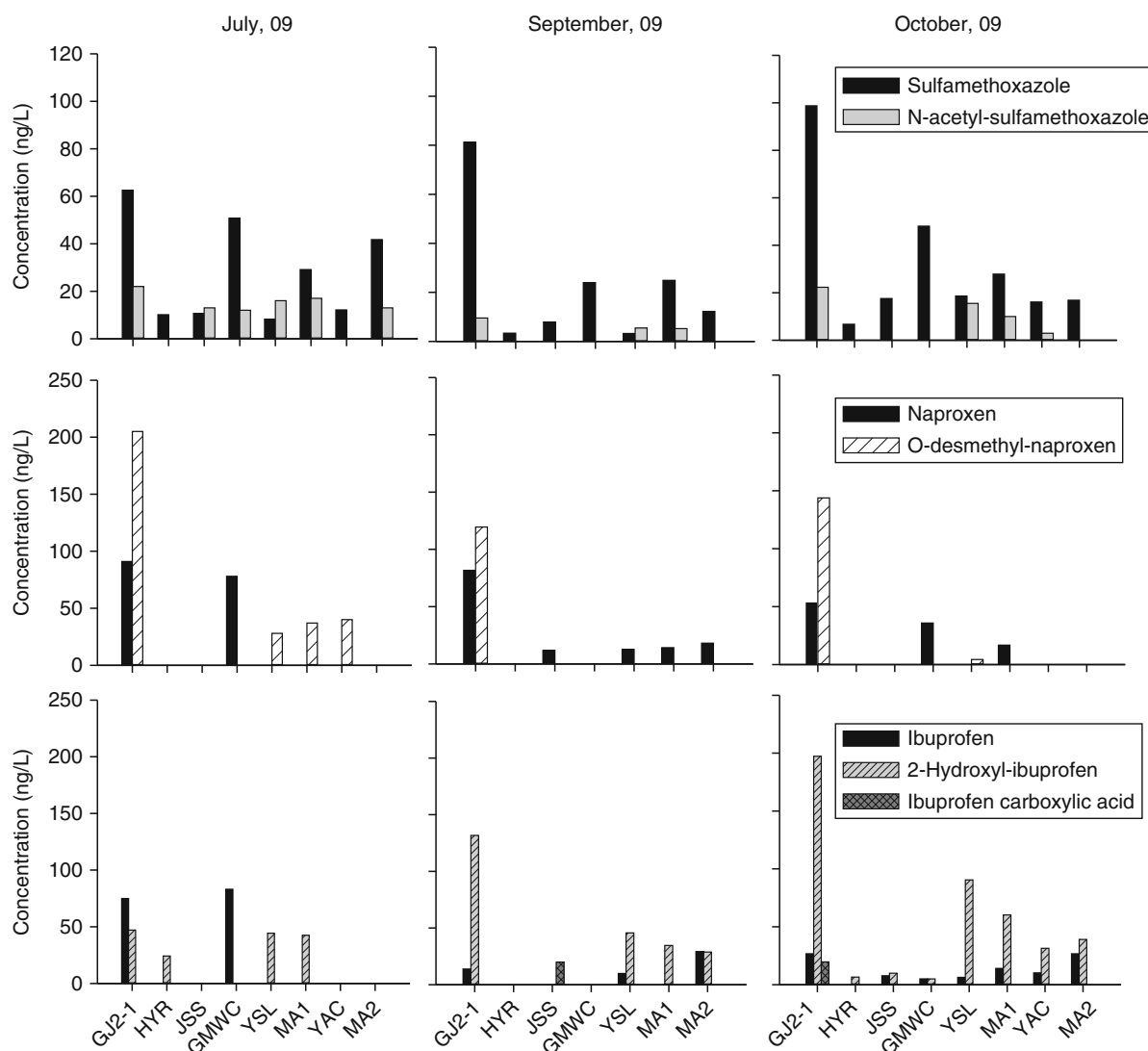
Spain. The concentrations of dilantin and atenolol were in good agreement with the results reported by Kim et al. [2] and Yoon et al. [10].

Occurrence of Pharmaceutical Metabolites in the River Water

As shown in Fig. 6, the occurrence of pharmaceutical metabolites was investigated along the Yeongsan River. A higher concentration of degradation products was observed in samples obtained from the GJ2-1 sampling site in Gwangju city (9.5–22.3 ng/L for *N*-acetylsulfamethoxazole, 119.4–205.0 ng/L for *O*-desmethyl-naproxen, and 47.1–197.4 ng/L for 2-hydroxyl-ibuprofen), whereas further downstream, the concentration of metabolites decreased, presumably due to dilution from natural attenuation. The prominent occurrence of pharmaceutical metabolites at the GJ2-1 sites is likely due to the effluent-dominated water source located in the Gwangju, which indicates that a high proportion of metabolites originated from human excretion in the WWTP influent and/or transformation of parent compounds into degradation products during WWTP processes mainly by activated sludge treatments. According to previous studies [33, 34], human excretion was believed to be the main source of ibuprofen metabolites. They also found that hydroxyl-ibuprofen was stable through WWTP processes and could be produced during activated sludge treatment, meanwhile carboxy-ibuprofen readily disappeared. This is in good agreement with the distribution of ibuprofen metabolites found in the Yeongsan River.

In previous studies [14, 35], a demethylated degradation product of naproxen, *O*-desmethyl-naproxen, was found in a lab-scale wastewater treatment unit. In this study, the ether cleavage and demethylation of naproxen were reported to be an initial step in bacterial degradation under aerobic conditions, as occurs in mammals. The naproxen metabolite observed in this study was also presumably transformed during both human metabolism and wastewater treatment processes.

In the case of the human metabolites of sulfamethoxazole, *N*-acetyl-sulfamethoxazole was found in surface water receiving WWTP effluent at a concentration range less than 50 ng/L [36], while the median concentration of this compound was 13 ng/L in this study. Another concern related to acetylsulfamethoxazole is



Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Figure 6 Spatial and seasonal concentrations ($n = 2$) of selected pharmaceutical metabolites and their parent compounds in the Yeongsan River, Korea

the retransformation back to the parent compound, sulfamethoxazole [37], implying reactivation of acetylsulfamethoxazole as an antibiotic compound during wastewater treatment, and/or after discharge into natural water.

Correlation Analysis Between Selected Water Quality Parameters and PPCPs

Correlation analysis was employed to evaluate the relationship between selected water quality parameters

(TN, DOC, UVA_{254nm} , SUVA) and 17 different target PPCP compounds, as shown in Table 4. TN, a representative wastewater characterization parameter, had the highest correlation with PPCPs; however, DOC, UVA_{254} , and SUVA were poorly correlated with PPCPs. This is inconsistent with previous studies [10, 20] that reported significant correlations between pharmaceuticals and organic matter parameters (TSS, DOC, COD, and BOD), which presumably resulted from the association between pharmaceuticals and

dissolved organic matter. In contrast to their explanation, the behavior of PPCPs in the Yeongsan River was not likely associated with dissolved organic matter.

Seven out of the 17 target compounds were found to be strongly correlated ($r > 7.0$) to TN, implying that these compounds could be potentially used as anthropogenic molecular tracers. The correlations were in the following order as determined from the regression coefficients obtained from the Pearson correlation at 95% confidence level (two-tailed); 2-hydroxyl-ibuprofen ($r = 0.87$) > atenolol ($r = 0.85$) > carbamazepine ($r = 0.81$) > sulfamethoxazole ($r = 0.79$) > dilantin ($r = 0.78$) > naproxen ($r = 0.76$) > iopromide ($r = 0.74$) (significant value < 0.000 for all). All seven compounds exhibited positive correlations with TN.

2-hydroxyl-ibuprofen was the highest correlated compound with TN, which indicates that hydroxyl-ibuprofen was much more stable and persistent in the wastewater effluent than its parent compound, ibuprofen, and the other wastewater-derived contaminants. According to Zwiener et al. [38], hydroxyl-ibuprofen was degraded much more slowly than ibuprofen in a biofilm reactor. This explanation is consistent with the results obtained and supports the notion that hydroxyl-ibuprofen is highly stable.

Atenolol, which was the second most highly correlated compound with TN, has been reported by Kasprzyk-Hordern et al. [16] as an anthropogenic indicator based on its persistent and ubiquitous occurrence. However, they also mentioned that atenolol could be susceptible to biodegradation through trickling filter and activated sludge treatments, indicating that it may be a suboptimal chemical indicator.

Carbamazepine has often been suggested as a conservative anthropogenic marker in aquatic environments [4, 7, 12, 16, 18]. The persistent behavior of carbamazepine corresponds to its low biodegradability (half-lives, $t_{1/2} > 100$ days) [3] and resistance to photodegradation. In this study, carbamazepine was highly correlated to TN with a regression coefficient of 0.81, which was consistent with other studies.

Another resistant compound in aquatic environments, sulfamethoxazole [11], had a relatively high correlation to TN ($r = 0.79$). Sulfamethoxazole has been described to be resistant to biotransformation with a half-life of > 100 days [3]. Since sulfamethoxazole is widely used as both a human and veterinary

antibiotic, it is difficult to determine if the presence of sulfamethoxazole at relatively high levels was caused by only human fecal contamination.

The anticonvulsant, dilantin, had a relatively high correlation with TN ($r = 0.78$), indicating that it may be a potential molecular marker. In a previous study [10], dilantin was shown to have a relatively high correlation ($r > 0.8$) with selected water quality parameters, including BOD₅, DOC, and UVA₂₅₄. In contrast, a negligible correlation between dilantin and selected water quality parameters, including DOC and UVA₂₅₄, was observed in the case of the Yeongsan River.

Anti-inflammatory, naproxen, which was highly correlated to TN with a regression coefficient of 0.76, has been regarded as a labile marker due to its photodegradability [4]. Other analgesics, such as ibuprofen and diclofenac, have been recommended as potential labile markers due to their susceptibility to biodegradation and photodegradation. Ibuprofen and diclofenac exhibited a lower correlation than the conservative markers such as carbamazepine and sulfamethoxazole, which had regression coefficients of 0.66 (p -value = 0.001) and 0.65 (p -value = 0.021), respectively. These results are in good agreement with previous studies [4, 16]. Kasprzyk-Hordern et al. [16] suggested that ibuprofen may be a potential indicator for non- or poorly treated wastewater, such as stormwater discharges, due to its readily biodegradable property in wastewater treatment.

X-ray contrast media, iopromide, can mainly be excreted without metabolism in the human body, and is rarely degraded in municipal WWTP [24]. With respect to its ubiquitous occurrence and high loads in surface water, and resistance to biodegradation, iopromide can serve as a potential tracer, which is supported by its high correlation coefficient to TN (0.74) (Table 4).

However, no significant correlations between caffeine, which has been previously considered a potential labile marker, and all the water quality parameters were observed in this study. Caffeine has been recommended as a labile chemical marker [17] with regard to its relatively high biodegradability (half-lives = 3.5–5.2 days) [3]. However, Young et al. [39] reported that caffeine was an inferior chemical indicator of sewage-derived microbes in surface waters with significantly low R^2 values between chemical and bacteriological indicators that ranged from 0.16 to 0.37, which was consistent

Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Table 4 Linear regression analysis between selected water quality parameters and PPCPs

Target micropollutants	n ^a	TN		DOC			UVA			SUVA		
		r ^b	F ^c	p ^d	r	F	p	r	F	p	r	p
ATN	30	0.85	72.3	0.000	0.34	3.76	0.063	0.22	1.42	0.244	-0.30	2.82
CPG	27	-0.34	3.29	0.082	-0.80	47.5	0.000	-0.89	91.5	0.000	0.27	1.96
DLT	25	0.78	36.0	0.000	0.59	12.3	0.002	0.67	18.7	0.000	-0.28	1.92
CBZ	31	0.81	53.8	0.000	0.44	6.98	0.013	0.39	5.19	0.030	-0.25	1.97
SMZ	32	0.79	49.7	0.000	0.29	2.72	0.109	0.24	1.88	0.181	-0.18	0.96
FFC	10	-0.47	2.28	0.170	-0.16	0.21	0.660	-0.11	0.10	0.760	0.13	0.14
NPX	19	0.76	23.0	0.000	0.62	10.6	0.005	0.71	17.5	0.001	-0.11	0.21
IBP	21	0.66	14.2	0.001	0.62	11.5	0.003	0.66	14.8	0.001	-0.31	2.00
DCF	12	0.65	7.45	0.021	0.40	1.91	0.197	0.33	1.19	0.302	-0.21	0.44
ACT	32	0.13	0.51	0.482	-0.26	2.20	0.149	-0.02	0.01	0.919	0.47	8.31
GMP	23	0.27	1.61	0.218	-0.08	0.14	0.715	0.33	2.60	0.122	0.60	11.9
CAF	32	0.22	1.54	0.224	-0.17	0.91	0.348	-0.14	0.56	0.461	0.22	1.48
IOP	30	0.74	33.3	0.000	0.18	0.89	0.353	0.05	0.08	0.780	-0.21	1.23
TCEP	31	0.001	0.0	0.996	-0.4	6.26	0.018	-0.47	8.23	0.008	0.19	1.09
N-acetyl-SMX	13	0.56	4.51	0.060	0.51	3.49	0.091	0.56	4.52	0.059	0.17	0.28
O-desmethyl-NPX	7	0.55	2.15	0.203	0.29	0.46	0.526	0.18	0.17	0.701	0.18	0.16
IBU-2OH	16	0.87	41.8	0.000	0.54	5.79	0.031	0.38	2.34	0.149	0.56	6.32

^aNumber of detected target compounds by LC-MS/MS

^bRegression coefficient

^cF-ratio

^dSignificant value. Correlations higher than 0.7 are significant at $p < 0.05$ (shown in bold)

with the results. The poorly correlated patterns of caffeine could be explained by its high biodegradability and ubiquitous discharge, resulting in distinct detections at high levels even further upstream and in tributaries, regardless of sampling dates and sites [7].

Although TCEP had the highest median concentration and was broadly distributed throughout the Yeongsan River, it was poorly correlated with all the

water quality parameters. This could be caused by its sporadic occurrence, irrespective of the wastewater effluent sources.

Correlation Among the Target PPCP Compounds

Correlation among the target PPCP compounds were evaluated as summarized in Table 5. The group of

Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Table 5
Correlation matrix among investigated target PPCP compounds

	ATN	CPG	DLT	CBZ	SMZ	FFC	NPX	IBP	
ATN	1.00	−0.14	0.59 ^a	0.84 ^a	0.82 ^a	−0.32	0.81 ^a	0.69 ^a	
CPG		1.00	−0.68 ^a	−0.28	−0.10	0.17	−0.71 ^a	−0.83 ^a	
DLT			1.00	0.71 ^a	0.65 ^a	−0.52	0.54 ^a	0.56 ^a	
CBZ				1.00	0.92 ^a	−0.18	0.77 ^a	0.61 ^a	
SMZ					1.00	−0.18	0.69 ^a	0.52 ^a	
FFC						1.00	−1.00	−1.00	
NPX							1.00	0.80 ^a	
								1.00	
	DCF	ACT	GMP	CAF	IOP	TCEP	N-acetyl-SMZ	O-desmethyl-NPX	IBU-2OH
ATN	0.57 ^a	0.24	0.12	0.49 ^a	0.81 ^a	0.48 ^a	0.43	0.56	0.77 ^a
CPG	−0.01	0.17	−0.29	0.06	0.01	0.36	−0.36	0.05	−0.10
DLT	0.67 ^a	−0.06	0.55 ^a	−0.06	0.61 ^a	−0.02	0.80 ^a	0.71	0.81 ^a
CBZ	0.45	0.12	0.33	0.48 ^a	0.89 ^a	0.19	0.38	0.58	0.84 ^a
SMZ	0.40	0.19	0.37	0.48 ^a	0.88 ^a	0.26	0.30	0.38	0.79 ^a
FFC	−0.02	0.20	0.08	0.01	−0.14	0.26	−0.40	−0.36	−0.71
NPX	0.43	0.36	0.43	0.81 ^a	0.62 ^a	0.51 ^a	0.35	0.91	0.53
IBP	−0.08	0.06	0.45	0.59 ^a	0.53 ^a	0.31	0.80 ^a	−0.25	0.48
DCF	1.00	0.42	−0.13	0.00	0.41	0.01	0.25	0.99 ^a	0.26
ACT		1.00	0.38	0.43 ^a	0.20	0.27	0.09	0.22	−0.26
GMP			1.00	0.31	0.21	0.03	0.53	−0.26	0.06
CAF				1.00	0.39 ^a	0.61 ^a	−0.25	−0.12	−0.53 ^a
IOP					1.00	0.30	0.46	0.24	0.75 ^a
TCEP						1.00	−0.46	−0.19	−0.37
N-acetyl-SMZ							1.00	0.19	0.45
O-desmethyl-NPX								1.00	0.18
IBU-2OH									1.00

Correlation higher than 0.70 are significant at $p < 0.05$ (shown in bold)

^aCorrelation is significant at the 0.05 level (two-tailed)

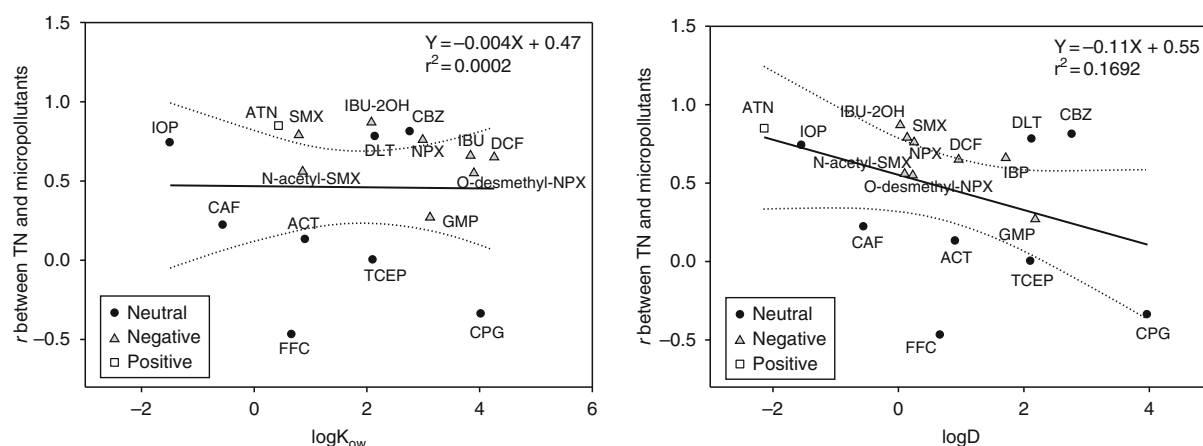
compounds with relatively high correlations to TN, including atenolol, carbamazepine, sulfamethoxazole, iopromide, and 2-hydroxyl-ibuprofen, was found to have significant correlations to each other with regression coefficients higher than 0.70 and p -values < 0.05 . This result indicates that these compounds may have similar behaviors in river water due to their conservative properties. Glassmeyer et al. [7] reported that carbamazepine and sulfamethoxazole were highly correlated due to their recalcitrant behaviors.

Naproxen exhibited high correlations with labile tracers, such as caffeine and ibuprofen. The high correlations among these compounds implied that they displayed an analogous mobility pattern in river water and hold promise for use as labile anthropogenic markers.

Some compounds, such as florfenicol, acetaminophen, glimepiride, and TCEP, were not correlated to other PPCPs, presumably due to their high biodegradability (for florfenicol and acetaminophen) and/or sporadic distribution (for TCEP and glimepiride). No distinct correlations between the parent compound and its metabolites were observed. Among the metabolites, 2-hydroxyl-ibuprofen showed higher correlation to other PPCPs than *N*-acetylsulfamethoxazole and *O*-desmethyl-naproxen, which was most likely due to the relatively high stability of 2-hydroxyl-ibuprofen compared to the other two metabolites.

Relationship Between Regression Coefficient (R) and Chemical Properties of Compounds

The pharmaceutical behavior was evaluated with respect to chemical properties (Fig. 7), such as acidity constant (pK_a) and octanol-water partitioning coefficient ($\log K_{ow}$ and $\log D$, i.e., $\log K_{ow}$ at pH7), by comparing the regression coefficient (r) obtained from the linear regression analysis between TN and target compounds (Table 4). As a result, $\log K_{ow}$ had no noticeable correlation to the regression coefficient (r), whereas $\log D$ had a slightly negative correlation to the regression coefficient (r). In general, $\log K_{ow}$ has been used to evaluate the pharmaceutical removal mechanism especially by sorption on activated carbon [40]. Lee et al. [21] used the $\log D$ parameter to elucidate the removal mechanism in a constructed wetland connected to wastewater treatment effluent, and found that $\log D$ was more suitable to predict the behavior of micropollutants than $\log K_{ow}$, which was consistent with the results obtained. The r^2 value obtained from the r - $\log D$ plot was too low to verify a linear relationship between r and $\log D$. However, by considering the complex removal mechanism including not only sorption but also biodegradation, oxidation, and photodegradation in river water, this low r^2 value indicates that sorption was not the main mechanism of



Wastewater Treatment Plant: Anthropogenic Micropollutant Indicators for Sustainable River Management. Figure 7 Relationship between regression coefficient (r) and octanol-water partitioning coefficient ($\log K_{ow}$ and $\log D$)

removal in the Yeongsan River. Despite this, the adverse proportional relationship between r and $\log D$ provides some information on the behavior of micropollutants in river water. Hydrophilic compounds with relatively low $\log D$ values had higher correlations to TN than hydrophobic compounds, indicating relatively low removal efficiencies for hydrophilic compounds. In addition, these results suggest that removal of hydrophobic compounds in river water occurs through sorption on organic matter. On the other hand, based on the pK_a values of target micropollutants, with respect to the functional groups of micropollutants at neutral pH, negatively charged micropollutants exhibited relatively high correlation to TN compared to the neutral charged compounds (Fig. 7). In summary, relatively hydrophilic and/or charged compounds, such as 2-hydroxyl-ibuprofen, sulfamethoxazole, naproxen, atenolol, and iopromide, play important roles as labile anthropogenic molecular markers in river water, with relatively low sorption tendencies and a relatively high biodegradation potential.

Conclusion

Eighteen different micropollutants including pharmaceuticals and their metabolites were investigated to evaluate their potential as anthropogenic molecular markers in Yeongsan River in Korea. Based on the spatial and temporal evaluation, most of the target compounds were frequently detected, exhibiting concentrations between 2 and 1,500 ng/L. In addition, these compounds were mainly discharged from the Gwangju metropolitan city, which was found to be the main source of the anthropogenic wastewater contaminants in the Yeongsan River ecosystem. The relationships between water quality parameters (TN, DOC, UVA, and SUVA) and micropollutants were examined by means of linear regression analysis. TN exhibited a strong correlation with seven compounds in the following order: 2-hydroxyl-ibuprofen, atenolol, carbamazepine, sulfamethoxazole, dilantin, naproxen, and iopromide. Through correlation analysis, the behavior of selected micropollutants was evaluated in terms of their use as conservative anthropogenic indicators (carbamazepine, sulfamethoxazole, iopromide, atenolol, and 2-hydroxyl-ibuprofen) and as labile wastewater indicators (naproxen, ibuprofen, caffeine, and diclofenac).

The $\log D$ and pK_a of each compound was then used to elucidate the chemico-physical behavior of target micropollutants as anthropogenic indicators in surface water. This analysis demonstrated that the $\log D$ and pK_a values could be used to assess the suitability of potential labile molecular indicators in river water. Even though there are still some limitations in understanding the detailed behavior of candidate indicators, the chemicals suggested as potential wastewater indicators in this study are expected to be helpful in analyzing sewage contamination in river ecosystems.

Future Directions

As growing concerns of the global climate change, the need for the development of technology for water treatment is also increasing to prepare for the water scarcity. Among various technologies, water reclamation and reuse is recommended for the alternative water treatments with high sustainability. In the sustainable water treatment, however, the existence of micropollutants is another concern. Even though there is no specific regulation related to the micropollutants yet, the removal of micropollutants in aquatic environment is considered to be important for reliable water sources in the manner of sustainability. Enhancement of water quality management system in terms of micropollutant control is a key step in the sustainable water treatment.

In addition, understanding the degradation pathway of micropollutants is also an important approach to effectively control the micropollutants in water bodies. Micropollutants are differently degraded into metabolites depending on their different physicochemical properties. Thus, removal efficiencies of micropollutants are differently affected by various removal mechanisms in terms of biodegradation, photodegradation, and oxidation/reduction. Research of micropollutants degradation pathway in aquatic environment would be helpful to understand behaviors of micropollutants and their metabolite and provide further research directions related to the removal of micropollutants.

Natural organic matter (NOM) is believed as a promising factor for evaluating behavior of micropollutants in surface waters. In previous study [41], NOM has been known to play an important role in photodegradation processes in surface waters. As radical scavenger and as precursor of reactive species such as singlet oxygen,

hydroxyl radicals, and others, NOM can promote the photochemical transformation of micropollutants. However, most of the role of NOM in the degradation of micropollutants is still unknown. Understanding the complex interactions between NOM and micropollutants is believed to provide interdisciplinary perspectives for sustainable water quality management.

Acknowledgments

This research was supported by the Yeongsan and Seomjin River Watershed Management Committee, Ministry of Environment, Republic of Korea.

Bibliography

Primary Literature

- Vanderford BJ, Snyder SA (2006) Analysis of pharmaceuticals in water by isotope dilution liquid chromatography/tandem mass spectrometry. *Environ Sci Technol* 40:7312–7320
- Kim SD, Cho J, Kim IS, Vanderford BJ, Snyder SA (2007) Occurrence and removal of pharmaceuticals and endocrine disruptors in South Korean surface, drinking, and waste waters. *Water Res* 41:1013–1021
- Benotti MJ, Brownawell BJ (2009) Microbial degradation of pharmaceuticals in estuarine and coastal seawater. *Environ Pollut* 157:994–1002
- Nakada N, Kiri K, Shinohara H, Harada A, Kuroda K, Takizawa S, Takada H (2008) Evaluation of pharmaceuticals and personal care products as water-soluble molecular markers of sewage. *Environ Sci Technol* 42:6347–6353
- Ternes TA (1998) Occurrence of drugs in German sewage treatment plants and rivers. *Water Res* 32:3245–3260
- Kolpin DW, Furlong ET, Meyer MT, Thurman EM, Zaugg SD, Barber LB, Buxton HT (2002) Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999–2000: a national reconnaissance. *Environ Sci Technol* 36:1202–1211
- Glassmeyer ST, Furlong ET, Kolpin DW, Cahill JD, Zaugg SD, Werner SL, Meyer MT, Kryak DD (2005) Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. *Environ Sci Technol* 39:5157–5169
- Fono LJ, Kolodziej EP, Sedlak DL (2006) Attenuation of wastewater-derived contaminants in an effluent-dominated river. *Environ Sci Technol* 40:7257–7262
- Wiegel S, Aulinger A, Brockmeyer R, Harms H, Löffler J, Reincke H, Schmidt R, Stachel B, von Tümpling W, Wanke A (2004) Pharmaceuticals in the river Elbe and its tributaries. *Chemosphere* 57:107–126
- Yoon Y, Ryu J, Oh J, Choi BG, Snyder SA (2010) Occurrence of endocrine disrupting compounds, pharmaceuticals, and personal care products in the Han River (Seoul, South Korea). *Sci Total Environ* 408:636–643
- Godfrey E, Woessner WW, Benotti MJ (2007) Pharmaceuticals in on-site sewage effluent and ground water, western Montana. *Ground Water* 45:263–271
- Benotti MJ, Brownawell BJ (2007) Distributions of pharmaceuticals in an urban estuary during both dry- and wet-weather conditions. *Environ Sci Technol* 41:5795–5802
- Benotti MJ, Trenholm RA, Vanderford BJ, Holady JC, Stanford BD, Snyder SA (2009) Pharmaceuticals and endocrine disrupting compounds in U.S. drinking water. *Environ Sci Technol* 43:597–603
- Kosjek T, Heath E, Kompare B (2007) Removal of pharmaceutical residues in a pilot wastewater treatment plant. *Anal Bioanal Chem* 387:1379–1387
- Onesios KM, Yu JT, Bouwer EJ (2009) Biodegradation and removal of pharmaceuticals and personal care products in treatment systems: a review. *Biodegradation* 20:441–466
- Kasprzyk-Hordern B, Dinsdale RM, Guwy AJ (2009) Illicit drugs and pharmaceuticals in the environment – Forensic applications of environmental data, Part 2: pharmaceuticals as chemical markers of faecal water contamination. *Environ Pollut* 157:1778–1786
- Buerge IJ, Poiger T, Müller MD, Buser H-R (2003) Caffeine, an anthropogenic marker for wastewater contamination of surface waters. *Environ Sci Technol* 37:691–700
- Clara M, Strenn B, Kreuzinger N (2004) Carbamazepine as a possible anthropogenic marker in the aquatic environment: investigations on the behaviour of Carbamazepine in wastewater treatment and during groundwater infiltration. *Water Res* 38:947–954
- Seiler RL, Zaugg SD, Thomas JM, Howcroft DL (1999) Caffeine and pharmaceuticals as indicators of waste water contamination in wells. *Ground Water* 37:405–410
- Santos JL, Aparicio I, Callejon M, Alonso E (2009) Occurrence of pharmaceutically active compounds during 1-year period in wastewaters from four wastewater treatment plants in Seville (Spain). *J Hazard Mater* 164:1509–1516
- Lee S, Kang SI, Lim JL, Huh YJ, Cho J (2010, submitted) Evaluating controllability of pharmaceuticals and metabolites, in biologically-engineered processes, using corresponding octanol-water partitioning coefficient with consideration of ionizable functional groups
- USEPA (1999) Protocol for EPA approval of new methods for organic and inorganic analytes in wastewater and drinking water. U.S. Environmental Protection Agency, Washington, DC
- Fernández C, González-Doncel M, Pro J, Carbonell G, Tarazona JV (2010) Occurrence of pharmaceutically active compounds in surface waters of the henares-jarama-tajo river system (madrid, spain) and a potential risk characterization. *Sci Total Environ* 408:543–551
- Ternes TA, Hirsch R (2000) Occurrence and behavior of X-ray contrast media in sewage facilities and the aquatic environment. *Environ Sci Technol* 34:2741–2748

25. Schulz M, Löffler D, Wagner M, Ternes TA (2008) Transformation of the x-ray contrast medium iopromide in soil and biological wastewater treatment. *Environ Sci Technol* 42:7207–7217
26. Herber T (2002) Occurrence, fate, and removal of pharmaceutical residues in the aquatic environment: a review of recent research data. *Toxicol Lett* 131:5–17
27. Winkler M, Lawrence JR, Neu TR (2001) Selective degradation of ibuprofen and clofibric acid in two model river biofilm systems. *Water Res* 35:3197–3205
28. Tixier C, Singer HP, Oellers S, Müller SR (2003) Occurrence and fate of carbamazepine, clofibric acid, diclofenac, ibuprofen, ketoprofen, and naproxen in surface waters. *Environ Sci Technol* 37:1061–1068
29. Buser H-R, Poiger T, Müller MD (1998) Occurrence and fate of the pharmaceutical drug diclofenac in surface waters: rapid photodegradation in a lake. *Environ Sci Technol* 32:3449–3456
30. Zwiener C (2007) Occurrence and analysis of pharmaceuticals and their transformation products in drinking water treatment. *Anal Bioanal Chem* 387:1159–1162
31. Halling-Sørensen B, Nielsen SN, Lanzky PF, Ingerslev F, Holten Lützhøft HC, Jørgensen SE (1998) Occurrence, fate and effects of pharmaceutical substances in the environment- a review. *Chemosphere* 36:357–393
32. Huerta-Fontela M, Galceran MT, Ventura F (2010) Fast liquid chromatography–quadrupole-linear ion trap mass spectrometry for the analysis of pharmaceuticals and hormones in water resources. *J Chromatogr A* 1217(25):4212–4222
33. Stumpf M, Ternes TA, Haberer K, Baumann W (1998) Isolation of ibuprofen-metabolites and their importance as pollutants of the aquatic environment. *Vom Wasser* 91:291–303
34. Weigel S, Berger U, Jensen E, Kallenborn R, Thoresen H, Hühnerfuss H (2004) Determination of selected pharmaceuticals and caffeine in sewage and seawater from Tromsø/Norway with emphasis on ibuprofen and its metabolites. *Chemosphere* 56:583–592
35. Quintana JB, Weiss S, Reemtsma T (2005) Pathways and metabolites of microbial degradation of selected acidic pharmaceutical and their occurrence in municipal wastewater treated by a membrane bioreactor. *Water Res* 39:2654–2664
36. Roberts PH, Thomas KV (2006) The occurrence of selected pharmaceuticals in wastewater effluent and surface waters of the lower Tyne catchment. *Sci Total Environ* 356:143–153
37. Göbel A, Thomsen A, McArdell CA, Joss A, Giger W (2005) Occurrence and sorption behavior of sulfonamides, macrolides, and trimethoprim in activated sludge treatment. *Environ Sci Technol* 39:3981–3989
38. Zwiener C, Seeger S, Glauner T, Frimmel FH (2002) Metabolites from the biodegradation of pharmaceutical residues of ibuprofen in biofilm reactors and batch experiments. *Anal Bioanal Chem* 372:569–575
39. Young TA, Heidler J, Matos-Pérez CR, Sapkota A, Toler T, Gibson KE, Schwab KJ, Halden RU (2008) Ab Initio and in Situ comparison of caffeine, triclosan, and triclocarban as indicators of sewage-derived microbes in surface waters. *Environ Sci Technol* 42:3335–3340
40. Westerhoff P, Yoon Y, Snyder S, Wert E (2005) Fate of endocrine-disruptor, pharmaceutical, and personal care product chemicals during simulated drinking water treatment processes. *Environ Sci Technol* 39:6649–6663
41. Doll TE, Frimmel FH (2003) Fate of pharmaceuticals-photodegradation by simulated solar UV-light. *Chemosphere* 52:1757–1769

Books and Reviews

- Asano T, Burton FL, Leverenz HL, Tsuchihashi R, Tchobanoglous G (2007) *Water reuse: issues, technologies, and applications*. McGraw-Hill, New York
- Niessen WMA (2006) *Liquid chromatography-mass spectrometry*, 3rd edn. Taylor & Francis, Boca Raton
- Snyder SA, Wert EC, Lei H, Westerhoff P, Yoon Y (2007) *Removal of EDCs and pharmaceuticals in drinking and reuse treatment processes*. Awwa Research Foundation, Denver

Water and Energy Nexus

GUSTAF OLSSON

Industrial Electrical Engineering and Automation (IEA), Lund University, Lund, Sweden

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Water and Energy Interdependencies
 Energy Use in Water and Wastewater Operations
 Energy on the Demand Side
 Energy Recovery from Wastewater
 Monitoring and Control for Better Efficiency
 Water Use and Impacts on the Aquatic Environment from Energy Production
 Conclusions
 Future Issues
 Bibliography

Glossary

Anaerobic Conditions in a biological treatments system characterized by the absence of oxygen in any of its forms.

Anoxic No oxygen present – nitrate instead of oxygen is used by the organisms.

Aquifer A natural underground layer that contains water.

BOD Biochemical oxygen demand, a measure of the organic carbon content in the wastewater.

Brackish water Water that is neither fresh nor salt.

Denitrification The conversion of nitrate-nitrogen to gaseous nitrogen through anoxic cell growth.

Desalination The changing of salt or brackish water into fresh water.

Eutrophication A significant increase in the concentration of chemical nutrients in an ecosystem.

Evaporation The process of liquid water becoming water vapour.

Fresh water Water that contains only small amounts of dissolved solids.

Groundwater Water that is pumped from aquifers.

Nitrification The conversion of ammonia-nitrogen to nitrite and nitrate-nitrogen through cell growth.

Renewable resources Total resources offered by the average annual natural inflow and runoff that feed a catchment area or aquifer natural resources that, after exploitation, can return to their previous stock levels by the natural processes of growth or replenishment.

Salt water Water that contains significant amounts of dissolved solids.

Surface water Water pumped from sources open to the atmosphere, such as rivers, lakes, and reservoirs.

Definition of the Subject

Water and energy are inextricably linked. Water is needed to generate energy and energy is needed to extract, treat, and distribute water and to clean the used and polluted water. This is the water–energy nexus and as a consequence, both challenges must be addressed together. Energy, water, and environmental sustainability are closely interrelated and are vital not only to the economy but to the health and welfare of all humans.

As a consequence of the close interrelationship between water and energy, the design and operation of water and wastewater systems should take the energy aspect into consideration. Similarly, energy production cannot be planned without taking water resources and water quality into consideration. Water availability is often undervalued and taken for granted. Population growth, climate change, urbanization, and rising health and environmental standards increasingly call for an

integrated approach. The design of our cities, suburbs, homes, and appliances has enormous implication for water and energy consumption. As a result, we cannot continue to utilize the critical resources water and energy in an inefficient and wasteful manner.

In this chapter, after an introductory section, we will focus on the following main issues:

- [Section 2](#): How the interrelationships between water and energy look like
- [Section 3](#): Efficient design and operation of water and wastewater facilities
- [Section 4](#): Attention to the water-related energy use at the end user
- [Section 5](#): Energy recovery from wastewater, such as biogas production
- [Section 6](#): The use of monitoring and control to make operations more efficient
- [Section 7](#): Water requirements for energy production

In the last two sections, [Conclusion](#) and [Future Issues](#), we summarize some of the findings and look for some future directions to deal with the water and energy nexus.

Introduction

Water has always been mankind's most precious resource – there are no substitutes. The struggle to control water resources has shaped human political and economic history. Population growth and economic development are driving a steadily increasing demand for new clean water supplies and it is well documented that lack of access to clean water has major health implications. Many see the water security as the key environmental issue of the twenty-first century.

Water and energy are the critical elements of sustainable economic development – without access to both of them, economies cannot grow, jobs cannot be created, and poor people cannot move out of poverty. On a global basis, neither water nor energy are in short supply. What is in short supply is energy and water at a price that people can afford to buy. We have to find better ways to use the water wisely, in our homes, in agriculture, and in industry. Using the water wisely includes producing the potable water and cleaning the wastewater using less energy. Pumping water, pressurizing water distribution systems, and pumping

wastewater are major energy consumers. Aeration of activated sludge systems is another major energy consumer.

Control and automation in water and wastewater operations can contribute to improve the sustainability by smart monitoring systems, disturbance attenuation, using energy wisely, and to maximize energy production.

Water and Energy Interdependencies

The interdependence between water and energy has been recognized during the last few years. Allan Hoffman wrote, in 2004 [1], on the topic: “The energy security of the United States is closely linked to the state of its water resources. No longer can water resources be taken for granted if the U.S. is to achieve energy security in the years and decades ahead. At the same time, U.S. water security cannot be guaranteed without careful attention to related energy issues. The two issues are inextricably linked.” In the USA, the energy–water nexus initiative was initiated in 2004 as an informal DOE National Laboratory initiative to develop a better understanding of the link between the nation’s energy and water supplies. The laboratories conducted preliminary assessments that indicated that the interdependence between energy and water supplies was much broader and much deeper than initially thought [1].

The increasing scarcity of water has been given much attention on the World Water Forums (1997 and every 3 years), UN Millenium Summit (2000), and now we are in the middle of the UN Decade of Water (2005–2015). Special workshops and conferences have been arranged, such as [2–4] and an interesting overview is found in [5].

Increasingly, there is a risk that water and energy interests are in conflict. Some examples can illustrate this:

- Between 1% and 18% of the electrical energy in urban areas is used to treat and transport water and wastewater. Furthermore, the energy related to water use – mostly heating the water in households and industries – requires about ten times more energy compared to the energy needed to deliver the clean and cold water and to treat the wastewater.
- To treat impaired water to drinking standards requires more energy. Most water treatment plants today are gravity draining through sand with some chlorine added. But as contaminants grow, old

technology does not work. Either membrane or thermal water treatment is more energy intensive than traditional methods.

- Thermal power plants require huge amounts of cooling water. For example, around 39% of all freshwater withdrawals in the USA are used for thermoelectric energy production. This is roughly the same amount of water as for irrigation [6]. Most of the cooling water is returned but around 3% is actually consumed, mostly by evaporation. Water availability has become a contentious siting issue for thermoelectric power plants and must compete with demands from municipalities, agriculture, and other industries.
- Climate change and increasing drought in many regions will decrease the flow in many rivers. Consequently, the cooling water may become insufficient for the need of the thermal power plants along the river.
- Energy exploration and production requires a lot of water and consequently will generate a lot of wastewater. Pure water is injected down a hole for oil or gas production, fracturing the shale formation to release oil or gas. Then, the water is brought back up with the oil or gas. The water is then severely contaminated. Today that water is injected thousands of meters down where it evaporates and is out of the water cycle.
- It is apparent that hydropower generation depends on water. A hydropower dam can be a serious environmental problem. The dam itself often serves as a gigantic sedimentation basin, and the solid material that earlier served as fertilizers downstream now are trapped in the dam. Obviously, the water flow downstream is affected. This has implications for the agriculture, water fauna, for recreations, and many other human activities. Increasing water shortage in combination with increased water use in many regions is now causing lack of water in the dams. With lower water levels, the generation of electricity is decreasing.

It is critical to understand the key roles and responsibilities associated with the management, operation, and use of the water and energy resources. This includes government agencies, private industry, as well as individual users. The interdependencies

between water and energy should force us to conduct planning and operation in such a way that both the water and energy flows are tracked. Such a systems analysis should recognize the inputs and outputs for each aspect of the system as well as the storage reservoirs for the entire system.

The Supply of Water

Considering the importance of the topic, it is interesting to examine how much resources have been spent on water resources research. The situation in the U.S. is interesting. Since 1973, the population of the United States has increased by 26%, the GDP and federal budget outlays have more than doubled, and federal funding for all research and development has almost doubled, while funding for water resources research has remained stagnant [7]. Still, the pressure on water resources increases with population and economic growth. There is an increase in the number of conflicts over water, so there are strong incentives to focus much more on the water and energy issues, both in many parts of the industrialized world and in the developing world.

The economy of water and energy behaves in a particular way [8]. When the supply is scarce, there is a high price inelasticity of demand. A small reduction of supply leads to a huge increase in price. As a result, the total value (price quantity) rapidly increases as total quantity declines [9]. This is true for any resource that is essential and non-substitutable. As there is less water or energy available, their price quickly increases. This can create havoc in markets and stress the whole economic system. Diminished water supplies may lead to direct conflict and violence.

When energy and water supplies are abundant, their value is low. It may seem that we have an infinite supply and there is no need to worry. However, as we approach depletion, even small perturbations due to unforeseen climatic events, sharp increases in demand, or technical malfunction result in disproportionate changes in their values and prices, if the market is allowed to work.

When energy supply of nonrenewable fossil fuels is depleted, we are looking for renewable sources. Similarly, we are over-utilizing fresh water resources in many places. This is compensated by increasing the

Water and Energy Nexus. Table 1 Energy requirements to pump 1 m³ to the water treatment in some different places (Data from [10, 11])

Location	kWh
Sweden (average)	0,22
Sweden lowest	0.04
Sweden highest	0.64
N. California, USA	0,04
S. California, USA	2,3
Melbourne, Australia	0,09
Adelaide, Australia	1,9

energy use to import water from other basins (such as S. California, Table 1), desalinate salt water, or reuse wastewater. Furthermore, nonrenewable water is depleted by over-pumping from fossil aquifers. These practices are not sustainable and will leave our children with fewer options. We need to think much more in terms of increased efficiency and conservation.

Energy Use in Water and Wastewater Operations

Energy is needed to extract water from underground aquifers, transport water through canals and pipes, manage and treat wastewater for reuse, and desalinate brackish and sea water to provide new fresh water sources. To get the energy data is not always easy. Statistics for water abstraction, treatment, and distributions are maintained by OECD and Eurostat and also by UN for different sectors. However, the corresponding energy consumption is usually not recorded in the water statistics, so it is not easy to recognize the interconnection.

Using the water wisely includes producing the potable water and cleaning the wastewater using less energy. Pumping water, treating the water to drinking quality, pressurizing water distribution systems, pumping and treating wastewater is a major energy consumer. In the USA and in the UK, this amounts to about 3% of the nation's electrical energy use. In Sweden, the corresponding number is 1%, while in Israel it is around 10%. This also means that water extraction and cleaning will have an impact on the greenhouse gas (GHG) emission. On a global basis, the water/wastewater

sector contributes with 1.5% of the CO₂ and about 5–7% of the GHG emissions. With increasing population and demand growth, the GHG emissions will most likely exceed 7% [9].

Some national water utility associations maintain statistics on energy consumption by the water sector. The power consumption follows typically a daily, weekly, and sometimes, seasonal use pattern. There is a potential for compensating some of the daily-use pattern by pumping to water storage reservoirs during low load hours for the power plants and vice versa for the high load hours.

Drinking water production is increasingly crucial, in particular in dry countries. The use of desalination is growing at a significant rate in many industrialized countries, and this development further emphasizes the close coupling between water and energy. At the same time, billions of people in developing countries have not yet turned on the first electrical light. This means that the structure of electrical energy generation and distribution has to be carefully considered. There is a wide variation of the electrical energy needed to produce drinking water. The large differences reflect both the different efforts needed to obtain the clean water and the water consumption. Table 1 shows some examples to illustrate the energy requirements to produce clean water.

S. California and Adelaide are dry places, and the water has to be pumped long distances. This is reflected in the very high energy costs for these places. The energy can be compared with the energy requirement to desalinate seawater, which is typically 4 kWh/m³. If there were unlimited energy, we would never have a problem with obtaining water for use. For example, there are vast resources of saline water that could be desalinated to provide for all the imaginable demands for water if there is energy to run those operations and then pump water to wherever it is needed. That is certainly not the case.

Wastewater collection, treatment and discharge, sludge treatment, and disposal require electrical power for pumping and aeration in biological treatment processes. Again, statistics for wastewater generation, collection, and treatment as well as sludge generation and disposal are maintained by OECD and Eurostat and also by UN for different sectors. As for water supply, the corresponding energy consumption is

seldom recorded in the water statistics. Some national water utility associations maintain statistics on energy consumption by the water sector.

Primary treatment of wastewater is sometimes called mechanical treatment. Pollutants are removed by sedimentation or filters and the solids are removed by scrapers. Biological nutrient removal (BNR) is a term used to describe plants that employ biological processes to remove organic matter (C) as well as nitrogen (N) and phosphorous (P). The activated sludge process was developed in the early 1900s with the purpose to remove components that consume oxygen. When the effect of eutrophication became known, the next important step involved the development of processes that could remove nitrogen and phosphorous.

The most important process for removing organic matter is biological oxidation, which involves microorganisms feeding on the carbon and the oxygen in the water. Some of the carbon is used for the growth of the microorganisms and the remainder is converted into carbon dioxide. In the microorganisms, around half of the organic matter is used to increase the body mass and half is converted into carbon dioxide.

Nitrogen principally arrives at the plant as ammonium (60–80%). Most nitrogen removal plants will transform the ammonium into free nitrogen that will escape via the water surface. The removal of nitrogen is a slower process than the removal of organic carbon and takes place in two principal stages, nitrification and denitrification. In the first process, ammonium is transformed into nitrate (an oxidation process) and in the second process, the nitrate is reduced to nitrogen gas.

The concentration of dissolved oxygen governs carbon removal, nitrification, as well as denitrification. In the carbon removal and nitrification, the process rate will increase with the oxygen concentration. However, there is a limit to the process rate, and higher dissolved oxygen concentrations will not help the biology but only waste energy for the compressors that aerate the biological reactor. On the other hand, with too little oxygen, the microorganisms (like humans!) will suffocate and the process rate will be significantly reduced. In the extreme case, the organisms will die. The opposite applies to the denitrification: the higher the level of dissolved oxygen the lower the rate.

To aerate the biological reactors is an energy-consuming process. This means that the aerator has to be controlled so that it balances between the biological need for oxygen and the energy cost to supply the air. Table 2 gives some Australian figures to illustrate the variability.

As a comparison, the UK water industry uses 0.63 kWh to treat 1 m³ of sewage and 0.59 kWh to treat 1 m³ of water [12].

The two major energy consumers in wastewater treatment are pumping of the influent wastewater, and aeration. The difference between the primary treatment and the biological and advanced treatment is mainly due to the aeration energy. Typically, aeration in the biological treatment systems represent around 50% of the wastewater operating costs, while pumping represents around 60% of water treatment operating costs.

To relate the energy use to the volume of water only tells part of the truth. In order to get adequate key performance indicators, it is important to relate the energy consumption both to the amount of organic components, nitrogen and phosphorus removed.

There is a significant difference between the energy requirements for different wastewater treatment plants. The differences are due to plant size, the type of load (e.g., industrial or mainly domestic), and the type of operation. For example, in Sweden, the energy requirement for wastewater operations varies from 1.5 to 40 kWh/kg BOD (organic carbon removal), with a median value of 4.5 kWh [11].

In some countries, there is already a demand to decrease the energy requirement for water and wastewater operations. In California, a 20% increase in

energy efficiency will be required according to the California Water Plan Update 2009. In China, the central government will require at least a 20% decrease in energy use [13] and in Sweden, a new energy savings program is being implemented where the goal is to save at least 20–30% in electrical energy requirement for wastewater treatment operations [11]. In Germany, the Ruhrverband has managed an ambitious energy savings program [14].

According to [2], it is quite feasible to obtain an energy consumption reduction by 20% by optimization and innovation. The existing systems in the water and wastewater industry have not reached the limits of improvement of its energy efficiency yet [15], [16]. It was also stated at the workshop [2] that a further reduction of the energy consumption with another 80% should be possible, but this requires a paradigm shift. The current water infrastructures have been designed and constructed on the basis of views, requirements, conditions, and technologies of decades ago. It is recognized that in the present systems, wastewater treatment, water treatment, and distribution are very energy intensive. New concepts could include topics like alternative sanitation approaches (vacuum system, separation at the source), from waste toward resource (P and N recovery; wastewater as nutrient for algal-based biofuel), microbial fuels cells, tailored water quality, and use of alternative resources, etc. The water and wastewater sector could benefit for technology developments and breakthrough in related areas like energy production, sensor development, nanotechnology, etc. Many research efforts have already been done in this direction [17].

Some of the important areas where energy efficiency needs to be addressed are:

- Water and wastewater pumping efficiency, including primary pumping, recirculations, and sludge pumping
- Motor controls, including soft start
- Aeration in biological treatment systems
- Mixing
- Water losses
- Demand side management – consumer behavior
- Operation of anaerobic digesters

Every energy efficiency improvement has to be tested from two angles: (1) How is the process affected,

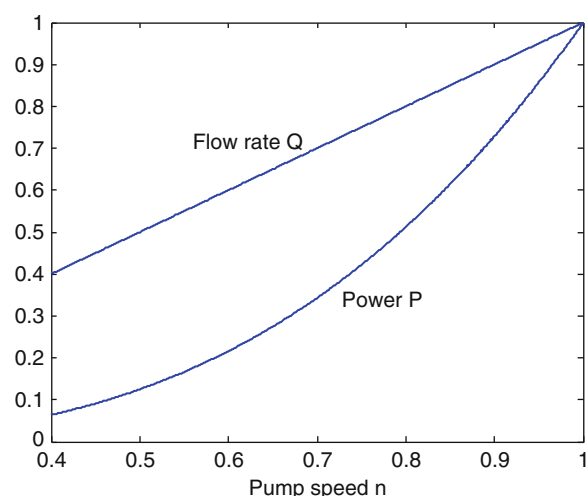
Water and Energy Nexus. Table 2 Energy requirement for wastewater treatment in Australia (Data from [10] and from Sydney Water and Brisbane Water)

Type of operation	kWh/m ³ (min, max)	kWh/m ³ average
Primary	0.1–0.37	0.22
Biological C removal (incl. primary)	0.26–0.82	0.46
Advanced C, N, and P removal	0.39–11	0.90

and (2) what is the cost of the efficiency improvement. Every wastewater treatment operator is responsible for the effluent water quality. Therefore, the regulatory requirements are always the primary goal and the energy efficiency becomes secondary. At the same time, it has been shown that sometimes a better efficiency will also create a better process performance.

There are two principal possibilities for better process efficiency, retrofitting the equipment and using control and automation. In the latter case, one also has to consider the investment in instrumentation, actuators (such as controllable pumps and compressors), and computer control systems.

Here, we will give a couple of examples of efficiency improvements by retrofitting or control. By designing pumps for adequate flows, operating at dynamically changing pressures and using variable speed pumping, there is a great potential for energy savings [18]. It is well-known that variable speed operation will save a lot of energy. Figure 1 illustrates the fact that the power requirement for pumping is proportional to n^3 , where n is the speed of the pump. At the same time, the flow rate is proportional to n . In other words, as the speed of the pump is halved, the flow rate will be halved while the power requirement decreases to 1/8. This is the very



Water and Energy Nexus. Figure 1

The relationship between the flow rate Q and the pump speed (n) (the upper curve) and the power P and the speed (the lower curve) according to the affinity laws

basic motivation to use variable speed drives in order to save energy.

It is not sufficient, however, to introduce variable speed pumping to save energy. The pump has to be designed so that its efficiency is at a maximum at the most common flow rate. This is a fact that is often overlooked in pumping of both water and wastewater systems.

Aeration of activated sludge systems is another major energy consumer. By automatic control of dissolved oxygen adjusted for variable wastewater loading, more energy savings are possible, and as a bonus, the biological activity is favored. There is a wealth of literature of dissolved oxygen control and a state-of-the art is given in [19], while [20–22] gives further experiences and evidence. In another chapter of this book [23], more details will be explained.

Often, the mixing of wastewater consumes too much energy. According to German experiences, 2–3 W/m³ is sufficient [14].

Energy savings can be obtained both by selecting the adequate equipment and by feedback control. Efficient pumping requires not only the proper pumps but also a control system that can adapt the pumping to the process needs. Likewise, energy efficient aeration requires not only sensors for dissolved oxygen and air flow rate but also controllable compressors and adequate air valves.

Water saving is directly related to energy saving, also in water industry operations. Leakage reduction, metering, recycling, and demand reduction all contribute to less waste of water.

Energy on the Demand Side

A lot of efforts are today spent on the supply side instead of on the conservation and efficiency side. For example, in Nevada, USA, the Southern Nevada Water Authority states [8]: “One of the main objectives of the Authority is to obtain additional water from Colorado River to support urban growth of its member agencies.” In general, hardly any water utility or water company has direct responsibility to reduce demand. At the same time, problems with energy and water supply are looming.

A lot of energy is used for bottled water. Only in the USA, the bottled water industry is a \$15 billion

industry and this industry consumes a lot of energy to manufacture, package, and transport bottled water over huge distances [8]. The average American in 1976 drank 6 l of bottled water a year, according to Beverage Marketing Corp. In 2009, the consumption was 108 l, more than 2 l/week. According to [24], “we are moving one billion bottles of water around a week in ships, trains, and trucks in the United States alone. That’s a weekly convoy equivalent to 37,800 18-wheelers delivering water.” Actually, 24% of the bottled water the Americans buy is tap water repackaged by Coke and Pepsi. Every year, 50 billion water bottles are used in the USA – 167 for each person – and 38 billion of them are pitched into landfills in the USA. This is more than \$1 billion worth of plastic that represents a lot of wasted energy. Actually, only 23% of the recyclable polyethylene terephthalate (PET) plastic bottles are actually recycled.

There is an interesting and strange attitude toward water. In a country like Sweden, the tap water quality is excellent, and from a quality point of view, bottled water will never be needed to replace the tap water. Still the Swedes consume around 30 l of bottled water per person and year. They pay 3–4 times the price for gasoline for the water, which is around 3,000 times the price for the tap water. In a recent examination (2009) of bottled water, only 2 out of 15 brands of bottled would be accepted as tap water according to the drinking quality standards.

Water usage has an energy cost. Both energy use and water use have to be sustainable. The use of suitable

energy sources is discussed elsewhere and is outside the scope of this chapter. However, also the consumption of water has to be sustainable. Recent droughts have put the focus on the water consumption in some countries. It is apparent that the water consumption is quite different in different countries. Looking at individual cities (100 cities compared), the specific water consumption varies globally from 0.34 to 650 l/capita/day [25], while the total charge for drinking water varies from \$0,015/m³ to \$3.13/m³. Some national comparisons are given in Table 3.

The Water Poverty Index has been developed in [27]. Canada, Norway, and Iceland are the countries with the highest index. No country has higher than 78. The index gives an assessment of the national water management and includes resources (water available), access to water supply and sanitation, economic standard, water use, and water environment.

It is true that the natural conditions are quite different. Still these figures often reflect the habits and sometimes misuse of water. Naturally, the price of water has a role to play here (see section “[Water and Energy Interdependencies](#)”). In general, the price paid by the consumer has a direct influence on the amount that is used. In the USA, the average household spends 0,006% of the income on water. The corresponding numbers for UK are 0,013%, for Pakistan 1.1%, and for Tanzania 5.7% [26]. It is most often true that the poor people pay more. While the average U.S. household pays around \$0.5/m³, and a German pays \$1.9, poor people often depend on informal vendors.

Water and Energy Nexus. Table 3 Water consumption in some countries [8, 26]

Country	Water use (l/capita/day)			Water poverty index (scale 0–100)
	Domestic	Agriculture	Industry	
USA	600	1,900	2,100	65
Australia	500	2,600	340	62
Canada	800	480	2,800	78
China	85	910	345	51
Germany	190	310	1,060	65
The Netherlands	80	460	820	69
India	140	1,500	95	53

A typical water price in Dhaka, Bangladesh is \$0.4/m³, in Phnom Penh, Cambodia \$1.6, and in Manila, Philippines \$4.7/m³ [26].

Curbing demand is cheaper, faster, and ultimately more beneficial to individuals than increasing supply. Conservation of energy saves both energy and water. Optimal water use saves both water and energy.

It should be realized that more than 90% of the water-related energy use is spent in the home. This is true for many industrialized countries. This fact is easy to realize just from a simple theoretical calculation. Usually, only a fraction of a kWh is needed to produce 1 m³ of drinking water. To heat this water from 15°C to 60°C requires 52 kWh, with no losses taken into consideration. Now, imagine how much warm water can be saved by using the showers, washing machines, dishwashers, etc. in a more economic way. Earlier it was stated that some 20–30% of the energy consumption for water and wastewater operations should be required. This saving should be profitable under most circumstances. Still, the same greenhouse gas emission savings could be achieved by only some 2–3% savings of warm water. Most water companies do not wish to take the responsibility for the domestic use of the water. Still, they usually have a good contact with the customers and should be able to easily influence the greenhouse gas emissions.

Energy Recovery from Wastewater

Energy recovery from sludge treatment is quite common and the biogas may be used for power generation, fuel for transport, or heating. In some cases, installation of heat pumps in wastewater treatment plant effluents has provided energy in terms of heat recovery. The power consumption in wastewater treatment and transportation follows typically a daily, weekly and sometimes, seasonal use pattern. Stormwater handling occasionally adds to the energy requirement. There is a potential for compensating some of the daily-use pattern by conducting certain treatment processes during low load hours for the power plants and vice versa for the high load hours.

A wastewater treatment plant is not only an “end-of-pipe” solution to clean wastewater. It can also serve as an energy producer. The interest in biogas is increasing and there is a large unused potential in maximizing

the production of biogas, used for both heating and as a replacement of gasoline in transportation, [28–31]. Recent data show that anaerobic digestion uses only some 20% of the energy content of the sewage. By-products from sewage treatment in combination with organic solid waste can provide a valuable source of energy if managed and utilized effectively. About one third of the chemical energy in the biogas can be transformed into mechanical energy. The rest of the energy becomes heat, but this can, of course, be recycled and used for heating. In order to use the biogas for engine fuel, the gas has to be further upgraded by removing carbon dioxide and hydrogen sulfide.

The biogas is produced in an anaerobic process where microorganisms break down the organic content of the sludge in the absence of oxygen. The process produces a methane and carbon dioxide-rich biogas. Furthermore, the nutrient-rich digestate can be used as fertilizer. The first part of the process is a so-called bacterial hydrolysis where organic polymers in the input material break down to become available as food for other organisms. Acidogenic bacteria then convert the sugars and amino acids into carbon dioxide, hydrogen, ammonia, and organic acids. These organic acids are then converted into acetic acid by Acetogenic bacteria, and additional ammonia, hydrogen, and carbon dioxide are generated. Finally, the products are converted to methane and carbon dioxide by the Methanogenic bacteria.

There are several factors that can influence the energy balance in the sludge handling in a wastewater treatment plant:

- *Primary sedimentation:* It may be possible to separate more sludge in the primary sedimentation. As a result, less organic carbon will be brought into the biological treatment step, so less aeration energy will be needed. In this choice, the organic carbon is chosen to produce biogas. However, organic carbon is also needed in the nitrogen removal process and insufficient carbon may limit the nitrogen removal capability.
- *Sludge thickening:* One way to increase the biogas production is to improve the thickening of sludge before the anaerobic digester. With a more efficient removal of water from the sludge, the sludge volume will decrease. Consequently, the retention time

of the digester can be increased. Also, less energy will be needed to heat the sludge before the digester.

- *Sludge heating*: The heat content of digested sludge can be used to heat the incoming sludge.

In cold countries, the heat content of the effluent water can provide significant heating not only to the plant facilities but also to city buildings via district heating systems [32]. In fact, most wastewater treatment plants can be net producers of energy.

There are other optimization aspects to consider. From an energy perspective, it looks straightforward to maximize the biogas production. However, from the digester, there will be reject water that has to be recycled to the wastewater treatment process. This water normally contains high concentrations of ammonium nitrogen. Therefore, the water that is recycled from the digester will significantly increase the load to the wastewater treatment plant, and thus to the energy requirement to operate the nitrification. This demonstrates that an energy balance has to be calculated in an integrated way and must consider the complete plant.

Monitoring and Control for Better Efficiency

Monitoring

To track the process operational state via the instrumentation is called *monitoring*. For the clean water supply, online monitoring will be required throughout the system, including at the tap. The availability of low-cost instrumentation will encourage better leakage detection [33, 34] and water quality monitoring [19]. In wastewater treatment systems, the use of instrumentation, control, and automation has proven to significantly reduce the costs for operation and detection, and early warning systems are key components in a sustainable water system. The development of reliable, affordable, and robust online instrumentation systems has been significant. Computing power no longer is a constraint, which means that the combination of measurements and model prediction can serve successfully in many systems.

Leakage detection in water distribution systems is becoming increasingly important as energy prices are raised. Leakages can appear as sudden bursts or they can appear as losses that will only gradually grow in size. Today, there are powerful automatic methods not

only to detect that a leakage or burst appears, but also to locate the position with great accuracy. This can be obtained both in single pipe systems and in pipe networks. The idea of early warning and problem detection is fundamental. The losses caused by late detection can be significant. Leakages cause not only losses to the various damages of the leaking water. The extra energy to pump leaking water is significant, and it is not unusual to find 30–40% water losses in water distribution system. This power loss is intolerable, but still surprisingly accepted.

Similarly, early warning systems composed by arrays of sensors in drinking water intakes can prevent severe quality disturbances into the system.

Automatic monitoring is also crucial in wastewater treatment operation. Of course, it should be related to the influent flow rate, concentration, and composition. Furthermore, monitoring equipment operation makes it possible to early detect malfunctions, so that major process faults can be avoided. Several examples of this are given in [19, 35–37].

Control

The fundamental principle of control is feedback. The process (e.g., an aerator, a chemical dosage system, or an anaerobic reactor) is all the time subject to disturbances. The current state of the process has to be measured by some sensor and this is the basis for a decision. In order to make a decision, the goal or purpose of has to be expressed. The controller calculates a decision that appears as an output signal from the controller. Via an actuator, which is typically a pump, a valve, or a compressor it will manipulate the process. In other words: *control is about how to operate the plant or process toward a defined goal, despite disturbances* [38].

Traditional wastewater treatment plant control is still unit process oriented to a great extent. Some examples of state-of-the-art control that will influence the energy efficiency are [19, 39]:

- Aeration control has the purpose to supply the microorganisms with adequate dissolved oxygen while saving maximum energy. Dissolved oxygen control will save a lot of electrical energy compared to no control at all. A time-varying setpoint of the dissolved oxygen concentration will further reduce the energy consumption.

- In a sequential batch reactor, the length of the different phases should be controlled. During the ammonia nitrogen oxidation time, the aeration is controlled until all the ammonia nitrogen has been consumed;
- The control of anaerobic processes aims at regulating the biogas flow, at stabilizing the process, and at maximizing its productivity. Still current state-of-the-art focuses on unit process operation;
- In section “[Energy Recovery from Wastewater](#)”, some examples of integrated control for biogas production were presented.

Water Use and Impacts on the Aquatic Environment from Energy Production

The demand for water is growing. The world water demand has more than tripled over the past half century. In the year 2000, the global water use was estimated to be about 30% of the world's total accessible fresh water supply. That fraction may reach 70% by 2025 [1].

This means that there will be more competition for the water needed for energy production. As a consequence, more water-intensive energy producers will be forced to look for more efficient alternatives. A key indicator may be “energy return on water invested” [8, 40].

Hydropower is often considered as “green” energy in the sense that the electric power production does not generate any emissions of greenhouse gases. Hydropower now generates about 19% of the world's electrical power. However, there are many signs of serious ecological impacts of hydropower generation. Dams do not only influence downstream transport of sediments, which has consequences for agriculture; they also serve as giant sedimentation basins. Fertilizers from the river no longer reach downstream agriculture. The sediments are instead collected on the bottom of the basin. Eutrophication in dams is another serious problem and evaporation from water dams in warm countries has not been sufficiently addressed as a great threat to freshwater availability. Dams can also act as physical barriers, restricting migration of aquatic wildlife.

On a global basis, only about one third of available hydropower potential has been used. However, the environmental and social costs for large dams will

probably prevent much further development of large hydropower systems. Still there is a potential for small hydropower plants – less than 10 MW – that could be adequate for remote rural areas. The reservoirs needed for large hydropower plants globally cover a surface area of around 500,000 km² – roughly the size of Spain. With a lot of hydropower in warm countries, there is a huge loss of water due to evaporation. The social cost is also significant. Around 80 million people have been forced to move due to dam constructions.

Thermal power plant operations – both coal fired and nuclear – require large amounts of cooling water. For example, in the USA almost 40% of the withdrawn freshwater is used for cooling in thermoelectric plants; however, only about 3% is actually consumed, mostly by evaporation. Heating of the water and the risk of radioactive contamination present great environmental challenges. Addition of chemicals, for example inhibitors, anti-scaling agents, and biocides in cooling towers, may represent high environmental loads of hazardous substances. Although these are at relatively low concentrations, accumulation of chemicals can be harmful in the long term. A wide variety of processes are already employed in power plants to recover, recycle, and reuse water. An overview of these technologies is found in [41].

Climate change will add to the risks of conflict between water and energy. The efficiency of the cooling process depends on the temperature difference between the cooling water and the temperature related to the plant. Consequently, there are strict requirements on the temperature of water that is used. With a higher cooling water temperature, more water is needed to provide the same cooling effect. Most plants have regulations or other constraints that limit their ability to adjust their withdrawal rates. In the short run, this means that they will get less cooling, a corresponding decrease in turbine backpressure, less efficient generation, and less electric energy for the same amount of raw energy input. Also, many nuclear plants have safety limits on intake temperature that could trigger complete shutdowns more frequently in altered climate scenarios. In addition, environmental concerns usually impose limitations on the temperature of water discharged back into the streams and reservoirs.

Water and energy industries are competing for the same resource. For instance, availability of cooling

water is a key issue for many power plants. If a river's flow is reduced, as they can do seasonally, power plants may find there isn't enough water available for cooling. This issue will only become more critical with climate change. It is quite apparent that the water quantity and quality aspects are not always placed sufficiently high on the agenda of many power companies. It will be increasingly important to establish networks of water and energy professionals to deal with the close relationship between water and energy.

It is obvious that energy savings in households and in industry will have a large impact on the need for electrical energy and consequently on water consumption and ecology. Technology development, for example, using more efficient pumps, refrigerators, air conditioners and heating systems, and power electronic motor control will offer a most significant impact on the total electrical energy use.

As an ironic consequence of technological development in developing countries, low-cost asynchronous motors and pumps have increased the water extraction for agriculture. One reason for the non-sustainable groundwater pumping is that the surface water is too contaminated. In some cases, this has had catastrophic consequences for groundwater levels and salinity of the soil. More efficient use of water for agriculture is therefore of paramount interest. This requires education, better equipment, probably economic support, and also some new legislation.

The "environmental friendly" zeal for ethanol has not considered the extensive irrigation required for the grain that supply ethanol. The situation in Midwestern USA can illustrate the problem. If irrigation is used, at least 1 m³ of water is needed for every 10 km travelled by an ethanol-powered vehicle [42]. According to [42], in case of ethanol-fuelled vehicles, there is a need to move from our old way of thinking – gasoline used per kilometer – to water used per kilometer. This water is often extracted from aquifers in a non-sustainable way. Fossil water is used and will not be replaced for generations.

It should also be noted that various portions of the ethanol production cycle have different water requirements. For example, in biofuel production irrigation requires orders of magnitude more water than ethanol biorefineries, as shown in Table 4. However, the intensity of water consumption can be much higher for

Water and Energy Nexus. Table 4 Estimated use of water for various technologies of biofuel production (From [8])

	Irrigation use (l water/l ethanol)	Refinery use
Oil		0.5–1 l water/l gasoline
Corn	0–1,900 l water	2–5 l water/l ethanol
Cellulose (sorghum) sugar	0–400 l water	6 l water/l ethanol

refineries, where thousands of cubic meters of water are to be withdrawn on the spot, significantly changing local hydrology and requiring additional infrastructure to provide that water.

According to estimates from U. S. Department of Energy (DOE), more than 4 l of water are needed for every liter of transportation fuel produced, threatening the limited water supplies (DOE Report 2006:80). China announced that it needs to curb coal-to-liquid production, because of concerns over pollution and the volumes of water consumed. Nevertheless, more recently, it was announced that the facility "will start operating later this year and is expected to convert 3.5 million tons of coal per year into 1 million tons of oil products such as diesel for cars." [8]. They will use groundwater and recycled water from coal mines to supply the 8 million tons it will need each year. In some parts of China, 30 years ago, the water table was 5 m below the ground. Today, it is 35–40 m below the ground because the groundwater is used in an unsustainable way.

Conclusions

Water and energy are the two most fundamental ingredients of modern civilization. Without water, people will die. Without energy, we cannot produce food or run our homes or industries. Considering the close relationship between water and energy, it is obvious that the challenges have to be treated in an integrated manner and single issues cannot be treated in isolation. An integrated approach means that the whole system has to be considered, including pure water resources, energy consumption, water usage, wastewater treatment, water

reuse, receiving water, and possible energy production. Integrated systems can only be considered by cooperation between several specialists. The interdisciplinary view has to be recognized, which also means that we have to exercise much more communication between engineers and scientists of different disciplines, but also between technology people and professionals in social sciences, behavior sciences, economy, and political decisions.

Saving energy and water is not just a technical challenge. Maybe the most important factor is our consumer attitudes and behavior. This includes how we use hot and cold water for showers, dishwashers, and washing machines. What kind of machinery is used? How do we consume water for gardening, etc.? As an example, the long drought in Queensland, Australia has forced the authorities in Brisbane to enforce a decreasing consumption. In 2005, the consumption was 300 L/day/person and in 2007, this had decreased to 130 [43]. This included restrictions for external water use and recommendations for internal water consumption.

As the world population grows, the demands for both water and energy increases faster than ever. It looks as if the world has become aware of the era of peak oil, seeing the fluctuations of the oil price. It seems that we are approaching an era of peak water – there will be lack of cheap water. The situation should already be considered a crisis, but the public in many developed countries has not grasped the urgency. Water is ultimately more important than oil, because it is more immediately crucial for life. We have government departments of energy. As Webber [42] points out, we should have a Department of Water that would ensure the effective use of water. Water is certainly a human right, but it should not be free or cheap. If we think that water is important, we should put a realistic price on it. The attitude toward water consumption may be the crucial ingredient. Furthermore, new approaches to financing, managing, and maintaining systems must be developed, as well as approaches to involve local communities.

The energy and water nexus has to be recognized by decision makers, researchers, and engineers as a vital one. It will not only determine the way to extract, treat, and distribute water, but also collect and treat wastewater. The role of water has to be fully recognized in the production of energy, both for electricity and for

transportation. For individuals, there are many ways to influence both the water and energy consumption in their daily lives, for example at home and when making transport choices. Creating the right pricing, policy, and regulatory environment is critical to encouraging behavioral changes, and ensuring a sustainable use of water and energy.

Conservation of energy saves both energy and water. Optimal water use saves both water and energy.

Future Issues

In the future, water supplies and treatment will probably become more energy intensive. Readily accessible freshwater supplies are limited or have been fully allocated in some areas. The new alternative water resources considered are sea water, brackish water, produced water, or impaired water. This means increased energy consumption for pumping at deeper depths and longer distances. New technologies to access and treat nontraditional water resources will require increased energy consumption per cubic meters of water.

Many new technologies offer interesting solutions. Wave energy offers sustainable solutions for energy generation, and the combination of desalination plants and wave energy generation is likely to be a very interesting area for research and development for water professionals in cooperation with power systems researchers. Microbial fuel cells (MFC) offer another interesting combination of water and energy issues. MFCs present an interesting energy challenge and also offer sustainable treatment of biological substances as a by-product [43].

Heat production from groundwater or effluent water in treatment plants is technologically possible, and application rates should increase along with investment in relevant infrastructure, for example, district heating systems could be applied at large scales.

Regulatory aspects of balancing the competing interests of water and energy needs by several sectors and industries is a multidisciplinary task. There is an emerging need to create professional platforms where such issues can be discussed and further developed on “neutral ground.” A focus should be on the development of concepts and motivating incentives rather than case-specific matters. For example, water pricing represents an issue not only affecting water and wastewater

utilities but also energy and agricultural sectors. It is obvious that new incentives and attitudes have to be developed in order to save water and save energy.

There is an urgent need for a paradigm shift from promoting growth to sustainability. Sustainable growth is impossible on a finite planet; we can only talk about sustainable development. We have created a global system that is closely coupled, and crises in one location send waves of disruption throughout the system. The fact that we live in a globalized system creates new opportunities, but also increases our risks, since there may be no place to go in case of a collapse. It will be unlikely that one developed country or region will be able to maintain its high quality of life if the rest of the world will be in substantial crisis. We have become very interdependent worldwide.

Bibliography

- Hoffman A (2004) The connection: water and energy security, Institute for the analysis of global security. <http://www.iags.org>, Aug 2004
- Water and energy (2008) Report of the GWRC research strategy workshop (www.globalwaterresearchcoalition.net), London
- IWA (2008) Water and energy workshop, IWA World Water Congress, Vienna, Austria. Presentations published on http://www.iwahq.org/Home/Themes/Water,_climate_and_energy/
- IWA (2009) Water & energy – mitigation in the water sector & potential synergies with the energy sector. In: International conference, international water association, Copenhagen, Denmark, 29–30 Oct 2009
- Pate R, Hightower M, Cameron C, Einfeld W (2007) Overview of energy-water interdependencies and the emerging energy demands on water resources. Sandia National Laboratories, Albuquerque, New Mexico, SAND 2007-1349 C. See also www.sandia.gov/energy-water
- Energy demands on water resources: report to congress on the Interdependencies of Energy and Water, DOE Report to congress, Jan 2007, <http://www.sandia.gov/energy-water/docs/121-RptToCongress-EWwEIAcomments-FINAL.pdf>. Accessed Feb 2010
- Webb EK, Johnson J (2009) Federal engagement in water resource technology development: current programs and the future. *J Contemp Water Res Educ* 143:3–16
- Voinov A, Cardwell H (2009) The energy-water nexus: why should we care? *J Contemp Water Res Educ* 143:17–29
- Farley J, Gaddis E (2007) An ecological economic assessment of restoration. In: Aronson J, Milton S, Blignaut J (eds) *Restoring natural capital: science, business and practice*. Island Press, Washington, DC
- Kenway SJ, Priestley A, Cook S, Seo S, Inman M, Gregory A, Hall M (2008) Energy use in the provision and consumption of urban water in Australia and New Zealand, CSIRO: Water for a Healthy Country National Research Flagship, CSIRO Australia and Water Services Association of Australia
- Lingsten A, Lundkvist M (2008) Description of the current energy use in water and wastewater systems in Sweden (in Swedish). The Swedish Water & Wastewater Association, SWWA, www.svensktvatten.se
- University of Oxford, Department of Earth Sciences, Environmental KTN, www.environmental-ktn.com. Accessed 9 Feb 2010
- Chen J (2008) Personal communication. Tsinghua University, Beijing
- Thöle D (2008) Ways to identify possibilities of energy saving at wastewater treatment plants. In: IWA (2008) Water and energy workshop, IWA World Water Congress, Vienna, Austria. Presentations published on http://www.iwahq.org/Home/Themes/Water,_climate_and_energy/
- Nowak O (2000) Expenditure on the operation of municipal wastewater treatment plants for nutrient removal. *Water Sci Technol* 41(9):281–289
- Alegre H, Baptista JM, Cabrera E Jr, Cubillo F, Duarte P, Hirner W, Merkel W, Parena R (2006) Performance indicators for water supply services, 2nd edn. IWA, London. ISBN 1843390515
- Urbanwater (2008) A Swedish research programme on “Sustainable Urban Water Management.” <http://www.urbanwater.org>
- Yates MA, Weybourne I (2001) Improving the energy efficiency of pumping systems. *J Water SRT Aqua* 50:101–111
- Olsson G, Nielsen MK, Yuan Z, Lynggaard-Jensen A, Steyer JP (2005) Instrumentation, control and automation in wastewater treatment systems. Scientific and Technical Report No.15, IWA, London, UK
- Serralta J, Ribes J, Seco A, Ferrer J (2002) A supervisory control system for optimising nitrogen removal and aeration energy consumption in wastewater treatment plants. *Water Sci Technol* 45(4–5):309–316
- Watts JB, Cavenor-Shaw AR (2008) Process audit and asset assessment using on-line instrumentation. *Water Sci Technol* 37(12):55–61
- Rosso D, Stenstrom MK, Larson LE (2008) Aeration of large-scale municipal wastewater treatment plants: state of the art. *Water Sci Technol* 57(7):973–978
- Olsson G (2010) Instrumentation, monitoring, control and automation in water and wastewater systems. IWA Publishers, London
- Fishman C (2007) Message in a bottle. *Fastcompany*. Dec 2007, <http://www.fastcompany.com/magazine/117/features-message-in-a-bottle.html>. Accessed Feb 2010
- International statistics for water services, International Water Association (IWA) Specialist group on statistics and Economics, IWA Biennial Conference, Vienna 2008
- Clarke R, King J (2006) *The atlas of water*. Earthscan, London
- Lawrence P, Meigh J, Sullivan C (2003) The water poverty index: an international comparison. *Keele Economic Research Papers* 2002/19, March 2003

28. Börjesson P (2008) Biogas from waste materials as transportation fuel—benefits from an environmental point of view. *Water Sci Technol* 57(2):271–275
29. Wiese J, Kujawski O (2008) Operational results of an agricultural biogas plant equipped with modern instrumentation and automation. *Water Sci Technol* 57(6):803–808
30. Siegrist H, Salzgeber D, Eugster J, Joss A (2008) Anammox brings WWTP closer to energy autarky due to increased biogas production and reduced aeration energy for N-removal. *Water Sci Technol* 57(3):383–388
31. Tilche A, Galatola M (2008) The potential of bio-methane as bio-fuel/bio-energy for reducing greenhouse gas emissions: a qualitative assessment for Europe in a life cycle perspective. *Water Sci Technol* 57(11):1683–1692
32. Fred, T. (2008). Large scale heat transfer from wastewater to city heating and cooling systems. In: IWA (2008) Water and energy workshop, IWA World Water Congress, Vienna, Austria. Presentations published on http://www.iwahq.org/Home/Themes/Water,_climate_and_energy/
33. Misiunas D, Lambert MF, Simpson AR, Olsson G (2005) Burst detection and Location in water distribution networks. *Water Sci Technol/Water Supply* 5(3–4):71–80
34. Misiunas D, Vitkovsky J, Olsson G, Simpson AR, Lambert MF (2005) Pipeline break detection using the transient monitoring. *J Water Resour Plann Manag ASCE* 131(4):316–325
35. Olsson G (2008) Process control. In: Henze M, van Loosdrecht M, Ekama G, Brdjanovic D (eds) *Biological wastewater treatment – principles, modelling and design*. IWA Publishers, London
36. Kaelin D, Rieger L, Eugster J, Rottermann K, Bänninger C, Siegrist H (2008) Potential of in-situ sensors with ion-selective electrodes for aeration control at wastewater treatment plants. *Water Sci Technol* 58(3):629–637
37. Shao-Yuan L, Rosso D, Jiang P, Larson LE, Stenstrom MK (2008) Real-time efficiency monitoring for wastewater aeration systems. *Water Pract Technol* 3:3
38. Olsson G, Newell B (1999) *Wastewater treatment systems. Modelling diagnosis and control*. IWA Publishers, London
39. Ingildsen P (2002). *Realising full-scale control in wastewater treatment systems using in situ nutrient sensors*. PhD thesis, Department of Industrial Automation, Lund University, Lund, Sweden (Access via www.iea.lth.se)
40. Spreng DT (1988) *Net energy requirements and the energy requirements of energy systems*. Praeger Press, New York
41. Wolfe JR, Goldstein RA, Maulbetsch JS, McGowin CR (2003) An electric power industry perspective on water use efficiency. *J Contemp Water Res Educ* 143:30–34
42. Webber M (2008) Catch-22: water versus energy, *Scientific American Earth* 3.0, pp 34–41
43. Keller J (2008) From microbial fuel cells to bio-electromechanical systems: how to convert organic pollutants to electric energy and more. In: IWA (2008) Water and energy workshop, IWA World Water Congress, Vienna, Austria. Presentations published on http://www.iwahq.org/Home/Themes/Water,_climate_and_energy/

Water and Wastewater Operation: Instrumentation, Monitoring, Control and Automation

GUSTAF OLSSON

Industrial Electrical Engineering and Automation (IEA), Lund University, Lund, Sweden

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Incentives for Control
 Basic Control Actions
 Control of Biological Oxidation
 Control of Denitrification
 Monitoring
 Plant Wide Control
 Future Directions
 Bibliography

Glossary

Actuator A transducer that reacts to a control signal and performs the desired action.

Anaerobic Conditions in a biological treatments system characterized by the absence of oxygen in any of its forms.

Anoxic Nitrate serves as terminal electron acceptor.

Closed loop control System where the controlled variable is measured and the result of the measurement used to manipulate one of the process variables. Also called feedback control.

Denitrification The conversion of nitrate nitrogen to gaseous nitrogen through anoxic heterotrophic cell growth.

DO Dissolved oxygen.

Nitrification The conversion of ammonia nitrogen to nitrite and nitrate nitrogen through autotrophic cell growth.

Online estimation Estimation of model parameters during the operation of the system.

Open loop control System where information about the controlled variable is not used to manipulate any of the system inputs to compensate for variations in the process variables. Also called sequencing.

ORP Oxidation-reduction potential or redox.

OUR Oxygen uptake rate is the rate at which oxygen is utilized by microorganisms, typically expressed in mg/L/h.

PID controller Three mode controller combining the actions of proportional (gain), integral (reset), and derivative (rate) elements into a single unit.

Redox Oxidation-reduction potential or ORP.

Setpoint The desired value for a control system, that is a temperature, flow rate, pressure, or level at which a process should operate.

WWTP Wastewater treatment plant.

Definition of the Subject

Instrumentation, control and automation (ICA) are now essential in all water and wastewater operations. While wastewater treatment plants (WWTPs) have been upgraded from relatively simple mechanical/biological plants for organic removal, the plants of today are increasingly complex and include not only biological nutrient removal but also the requirement to interact on both a plant wide level and with the sewer systems.

The ideal ICA system contains four functional components:

- A quality team of people who feel a deep sense of ownership of the system and the treatment plant and who are committed to the continuous improvement ethics.
- An instrumentation system to gather adequate process variable information.
- A monitoring system to acquire data, process and display the data, detect and isolate abnormal situations, assist in diagnosis and advice, and finally simulate the consequences of operational adjustments. A proper data acquisition and reporting is crucial.
- A control system to meet the goals of the operation. This can take place both locally within the treatment process by low-level control system or by coordination of the various processes within the plant as well as with the sewer system.

Automation in wastewater treatment systems have to include several disciplines:

- **Process engineering:** The focus of the ICA efforts always has to be the process. Therefore process knowledge, including the equipment to operate the process, is essential for a successful control. Dynamic modeling and simulation is used as an important tool.
- **Instrumentation technology:** Measurements of the states and parameters of the process requires instrumentation knowledge.
- **Signal processing:** Handling the measurement data has the purpose to remove inadequate data and extract essential information from the sensors and instruments.
- **Monitoring:** Based on the measurement information, a monitoring and detection function has to help to make the operation reliable.
- **Operator guidance:** Based on measurements and monitoring, the computer system should support the operator to make the adequate decisions.
- **Automatic control:** The controller makes the decision how to change the control variables (liquid and airflow rates, dosages, reaction times, etc.).
- **Actuator technology:** The design and operation of pumps, compressors, and valves.
- **Production planning:** This includes planning how to operate the plant to prepare for future disturbances and external changes. This can include planning for a coming rain event, extreme loads (like a tourist season), or temperature due to seasonal variations.

The concept of control is further described in the [Introduction](#). There are several incentives for control of wastewater treatment systems, and they are illustrated in [Incentives for Control](#). Disturbances to the plant motivate feedback control and this requires that the goal for the control is clearly defined. [Basic Control Actions](#) presents the basic ideas of control concepts and some of the basic control actions that have to take place to just keep the plant running. In [Control of Biological Oxidation](#), some specific control systems are described for the biological oxidation processes, where dissolved oxygen (DO) control is a key operation. [Control of Denitrification](#) describes how control can make the

reduction process denitrification more efficient and economical. Monitoring is to keep track of the current process operational state via the instrumentation. This is further described in [Monitoring](#). In order to satisfy increasing demands on effluent quality and ecological standards, it is needed to apply an integrated view of the whole plant operation and its connection upstream to the sewer network and downstream to the receiving water. These aspects are discussed in [Plant Wide Control](#). In [Future Directions](#), finally, some future perspectives are discussed.

Introduction

Advanced control is getting increasingly demanded in water and wastewater treatment systems and has been subject to a lot of applications in other parts of the process industry. Various case studies have shown significant savings in operating costs and remarkably short payback times. Process control in wastewater treatment systems is today well recognized as an integrated part of the plant operation. Still, however, the industry has a lot to learn from other process industries, like the chemical or paper-pulp process industries. It is the process knowledge, the sensor technology, and the way the plants have been designed and built that may limit what can be achieved today. Wastewater processes do have some unique features: the flow rates, the disturbances, the small concentrations, the organisms, the separation, and the fact that all the “raw material” has to be accepted and treated.

During many years, the incentives to apply automatic control were missing. Operators and organizations did not have the incentives to develop operation based on instrumentation, control and automation (ICA). Economic incentives are of course essential and today many organizations have created such economic driving forces that have accelerated the application of ICA.

There are several technical driving forces that have contributed to make ICA being applied much more:

- *Instrumentation development*: Today online instrumentation for some key components are available and getting both reliable and profitable to use. This includes sensors for ammonia, nitrite/nitrate, and phosphate.
- *Actuators*: Variable speed technology is well proven for electrical motors. This means that both pumps and compressors can be operated flexibly, which will have a significant impact on the quality of the operations.
- *Computer systems*: Distributed computer systems, including PLCs (programmable logical controllers) are available in modules and can be adjusted to the size of the plant.
- *Automatic control*: The control technology and engineering is sufficient for most needs and is often packaged in the real-time computer systems.
- *Process knowledge*: This has increased tremendously over the last decades. Dynamic models of the treatment systems are available in commercial simulators and this can be of help not only for the design but also for operator guidance.
- *Education*: The level of education of operators is often high and increasing. However, unfortunately the education of process designers most often does not include anything about dynamics and control. This was recognized already in 1974 at a conference at Clemson University, South Carolina, USA. Professor John Andrews recognized the need for education at all levels and he noted: “A course in *Process Dynamics and Control* is commonly found in most chemical engineering curricula. We would be well advised to include a course in *Dynamics and Control of Wastewater Treatment Systems* in environmental engineering curricula.” Today there are such courses at a few universities, and it has taken a long time to make this vision come true.

Demands from the regulatory agencies are also important driving forces and give incentives for control. Demands on organic matter content, nitrogen, and phosphorous components in the effluent require good operation consistently, around the clock. The only reasonable solution to this challenge is online measurements and automatic control.

Economic incentives are today much more apparent, and the requirement for economic operation are expressed clearly at many plants.

It is generally acknowledged that the introduction of ICA during the last decades has increased the capacity of biological nutrient removal (BNR) wastewater treatment plants by 10% to 30%. If these possibilities

are further exploited by intelligent use of measurements and information technology, the improvements due to ICA will have a significant impact during the next decade. It has the potential to reduce operating costs – such as electrical energy, chemicals, sludge disposal, and in some countries green taxes – and to keep the effluent quality consistently acceptable, despite load variability and plant disturbances. Furthermore, ICA can make the plant meet increasing load demands. For apparent reasons, ICA has to be integrated already in the plant design process.

As the complexity of the plants and processes has grown, the electrical and real-time computer systems have developed even faster. It is expected that ICA will have an even larger impact on both the performance and economy of wastewater treatment systems in the coming decade.

More detailed descriptions of the state of the art of ICA in wastewater treatment systems are given in the books [1–3].

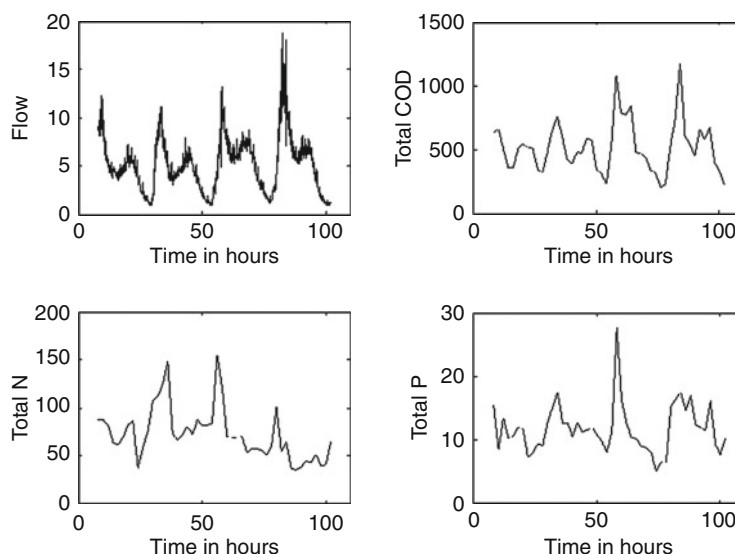
Incentives for Control

Control is about how to operate the plant or process toward a defined goal, despite disturbances.

Disturbances

Disturbances are the main reason why control is needed. They are present all the time and will affect the process, so they cannot be neglected. The load varies every minute, hour, day, and month with respect to its flow rate, its concentrations, and its composition. Under all these varying load conditions, the plant has to operate in such a way that it satisfies the effluent requirements. A wastewater treatment process is quite unique in the process industry. The “raw material” cannot be controlled and simply has to be accepted. Despite this lack of quality control of the input the “product,” the effluent water, has to satisfy a sufficiently good standard all around the clock.

Disturbances related to the influent can be described as in Fig. 1, which represents a municipal treatment plant in a suburb with no industrial plant connected to the wastewater treatment plant. Diurnal variations are apparent and the flow rate varies at least by a factor of 10. The concentration variations are almost in phase with the flow rate variation, so the total load varies significantly. Under all these conditions, the plant has to be operated satisfactorily. Traditional design usually solves the variability problem by increasing the tank



Water and Wastewater Operation: Instrumentation, Monitoring, Control and Automation. Figure 1

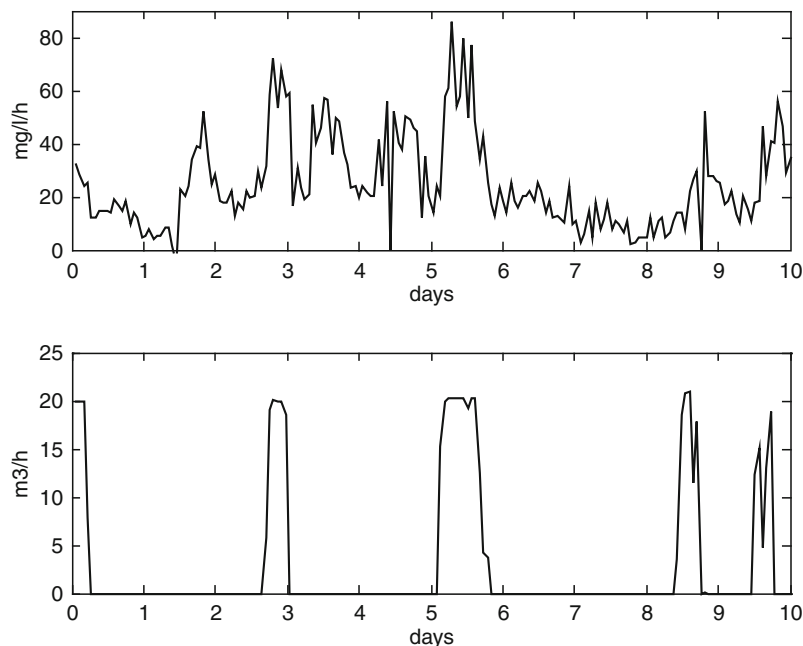
Diurnal variations from Thursday through Sunday in a municipality without any industry. There is no rain during this period

volumes. In this way, the variations are smoothed and the plant can produce a satisfactory effluent. However, the price of such an overdesign is high. Also, at low loads the plant may be “underloaded” and the retention time in the system will simply be too long.

Many treatment plants have to receive stormwater from rain. This causes two types of disturbances. One is of course the higher flow rate, which will decrease the retention time in the process. Furthermore, it may risk the performance of the settler, which will result in loss of sludge into the effluent. Thus, the effluent quality will deteriorate while the process will lose sludge-containing microorganisms that serve in the waste removal process. The other rain disturbance is due to the rate of change of the flow rate. A sudden rate of change will again cause problems in the settler unit. Any sedimentation has to guarantee a smooth laminar flow, but the sudden flow increase will propagate to the settler and cause hydraulic upsets and some turbulence. This is of course detrimental for the settler performance.

In many plants, there are unintentional disturbances caused by the equipment. If the plant is supplied with only on/off control of the influent pumps, any pump start will cause a sudden flow increase into the plant. The result is again an upset of the settler performance. Consequently, a good control and operation of the plant will assume smooth pumping using variable speed control of the pumps.

A wastewater treatment plant consists of many unit operations that are connected serially or via recirculations. As a result, a disturbance in one of the units will propagate to some other unit. These kinds of disturbances are less obvious, but may significantly influence the plant operation. One example can illustrate this. Reject water (supernatant) from an anaerobic digester is usually recirculated to the influent of the wastewater treatment plant. The supernatant has often a high concentration of nitrogen components and will consequently increase the load to the plant quite significantly. This can be illustrated by Fig. 2.



Water and Wastewater Operation: Instrumentation, Monitoring, Control and Automation. Figure 2

The consequence of recirculation of reject water (supernatant) from an anaerobic digester back to the treatment plant during a 10-day period. The lower curve shows the supernatant flow rate. The concentration is very high, so the impact on the wastewater treatment is significant. The upper curve shows the oxygen uptake rate in the aerator, which is a sign of the biological activity as well as the load (M.K. Nielsen, Denmark, personal communication, 2006)

If the supernatant is recirculated at a high load situation, the operation of the biological reactor will suffer. Therefore, it is important not only to control each one of the unit processes but also their interaction. This recirculation should be activated only at the low load periods.

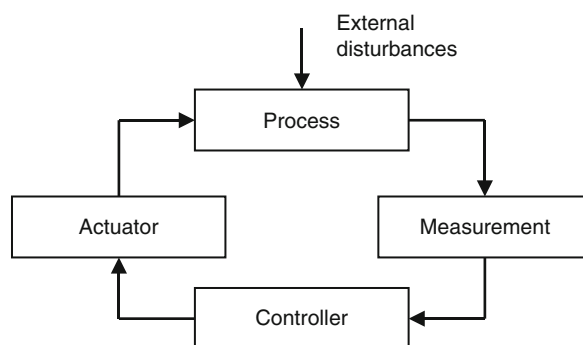
The Control Goals

The ultimate goal of the operation of a wastewater treatment plant is to satisfy the effluent criteria. They can be expressed in terms of organic material content, total nitrogen, total phosphorus, and suspended solids. However, given the many steps of the operations from the influent to the effluent it is most often almost impossible to base any control action on measurements of the effluent quality. Instead, a number of intermediate goals have to be formulated in order to synthesize some control action. Examples of such goals are:

- Grow the right biomass population
- Maintain adequate mixing
- Keep the level of dissolved oxygen at the right level
- Maintain an adequate airflow rate while minimizing the pressure
- Keep the sludge blanket level in the settler between given limits
- Avoid overload of the clarifier
- Avoid too long retention time of the sludge in the settler

Another goal is to *minimize the operational costs*. Here a control system can make a significant contribution to savings. The two major energy consumers in wastewater treatment are pumping of the influent wastewater and aeration. Typically, aeration in the biological treatment systems represents over 80% of the energy costs for biological treatment (Olsson G, ► [Water and Energy Nexus](#), this volume). Depending on the geographical conditions pumping costs can be quite different in different plants, but pumping together with aeration represents most of the energy cost.

Aeration can be controlled so that the microorganisms will receive just enough air to function, but not more. Any excess of air flow will cost money. Likewise, the pumps have to be designed in such



Water and Wastewater Operation: Instrumentation, Monitoring, Control and Automation. Figure 3
Illustration of the feedback principle

a way that they have a maximum efficiency at the most common flow rates. On top of this it is important to control the flow rates smoothly, so variable speed drives are important for good control.

Basic Control Actions

The fundamental principle of control is feedback, illustrated by Fig. 3. The state of the process is measured by some online sensor (or a human observer). Based on the measurement a decision is made how to correct the process in order to reach the goal. The decision is made in the controller that will send a command to some actuator that will make the correction of the process. The reason that measurements and corrections have to be made all the time is that disturbances will influence the process behavior in some way.

One is faced with feedback systems all the time in one's daily life. The human body would not function without a huge number of feedback loops. For example, to stand up requires feedback. The balance system in the ears and the vision will measure if the body is straight. The brain calculates how to change the feet in order to keep the balance and the muscles (the "actuators") will make the change happen.

While driving a car the driver all the time applies feedback. The eyes watch the speedometer and the road, etc., and the brain will compose all that

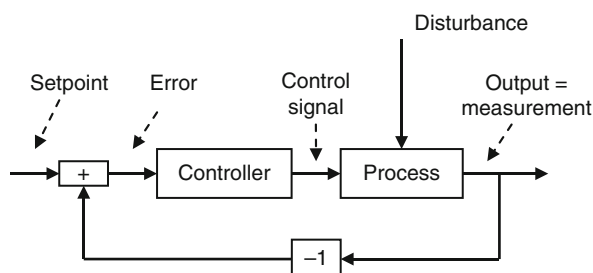
information and make a decision what to do in the next moment. This is translated to the muscles to turn the steering wheel, to brake, or to accelerate. The reason all the time for the feedback is that the scene is changing continuously. In other words, the “process” is subject to disturbances that force one to use feedback.

Open Loop Control

It is important to understand the difference between feedback control (also called closed loop control) and open loop control. In open loop control, the control action is simply triggered by a timer. No measurement of the result is made. Some common examples of open loop control are:

- A pump that is switched on certain time intervals and switched off the rest of the time. No flow rate measurement is made.
- A sequential batch reactor (SBR), where the phase length is only determined by a timer. For example, in nitrogen removal the reactor is aerated a given time and then not aerated in the next time interval. Again no measurement is made of the ammonium or nitrate concentrations at the end of each phase. More of this is discussed in [Control of Biological Oxidation](#) and [Control of Denitrification](#).
- A sludge scraper is operated every n th minute. The sludge level is not measured, so sometimes the scraper will remove only small amounts of sludge and at other times too much sludge.
- Two compressors may provide aeration for the biological reactor. One of them may be switched on during the day and switched off during the night. No confirmation of the resulting dissolved oxygen concentration is made.

It is obvious, that measurements and feedback control are necessary to truly compensate for variable loads and other disturbances. Open loop control may seem to have a lower cost, since it does not require measurements. However, if operating costs are calculated, then the open loop control is mostly inferior. Furthermore, if the effluent quality is to be maintained, then feedback is superior.



Water and Wastewater Operation: Instrumentation, Monitoring, Control and Automation. Figure 4
A simple control loop

Control of the Equipment

The top priority for any operation is to keep the plant running. This means that all the equipment, such as pumps, valves, motors, compressors, sludge scrapers, etc., simply have to work as intended. Many of these operations are not directly influencing the effluent quality, but without a proper operation of the equipment the effluent quality will suffer sooner or later.

Figure 4 is a control engineering block diagram and illustrates the principle of a simple control loop for the local controllers. Assume that one wishes to control the airflow rate in an aerator. The desired value (setpoint) may be provided in two different ways. The operator can manually set the desired airflow rate or another controller can send a signal that the airflow rate has to be changed. The desired value is compared with the measured airflow rate that can be measured for example by a differential pressure sensor. The controller acts on the difference between the desired and the measured values, the error. The controller tries to bring the error to zero in the best possible way. The controller will send its control signal to a valve and some motor or pneumatic device may open or close the air valve. The external disturbance in this case may be the varying oxygen demand from the microorganisms in the aerator.

Opening or closing an airflow valve means that the pressure in the air pipe will change. A pressure controller is used to keep the pressure constant, despite the disturbances. Again, the structure of the pressure controller looks like Fig. 4. The controller calculates the

difference between the desired pressure and the measured pressure and produces a control signal that will influence the airflow production from a compressor. This means that the pressure controller provides a new setpoint to the compressor.

It is assumed that the compressor has a variable speed drive system. Then the pressure controller has given the setpoint of the speed. The pressure controller will send a signal “increase the speed” in order to increase the pressure and this implicates that the speed controller setpoint is increased. The speed controller of a motor again looks similar to Fig. 4. Basically, the speed controller compares the actual motor speed with the desired speed and attempts to make the difference (error) equal to zero. Still, if the pressure sensor is not satisfied, then the speed setpoint will be further changed until the speed of the compressor is such that the pressure setpoint has been reached.

It is obvious that any wastewater treatment plant contains a large number of local controllers for physical variables like pressures, speeds, levels, liquid flow rates, and airflow rates. It is not at all apparent how each one of them will influence the final product, but all of them have to work satisfactorily in a well-functioning plant.

The control problems at the equipment level are very similar in any process industry, and a wastewater treatment plant is no exception. Almost all these controllers are of PI (proportional + integral) type, that is, the control signal is proportional to the error signal and to its time integral. Such a controller has the form:

$$u = u_0 + K \left(e + \frac{1}{T} \int e \, dt \right)$$

where u is the control signal, u_0 an offset signal, e the error, K the proportional gain, and T the integral time. In a local controller, the tuning of the two parameters K and T is the primary task. A too small K will make the controller response slow while a too large K will make the controller react too quickly. It may also cause instability of the system. The integral part of the controller has one purpose, to finally make the error equal to zero.

The design of feedback controllers has attracted considerable attention in the control literature. Many

advanced control algorithms based on for example dynamic models, neural networks, and fuzzy logics have been proposed. However, no convincing evidence has been made available suggesting that these advanced algorithms produce better control performance in wastewater treatment systems than the conventional PID (proportional-integral-derivative) algorithms, which have been used in most practical process control applications (more than 95% of the controllers in a typical paper and pulp industry are PID controllers). Control systems based on simple rules (rule-based control) have also found successful applications.

All measurements on the equipment level are physical measurements, such as levels, pressures, air and liquid flow rates, so the sensor technology is well established [4]. There are literally hundreds of textbooks in control, so just two are selected for the bibliography. The textbooks [5], [6] give a good introduction into control theory applied for process control.

Control of Biological Oxidation

The control actions in the various unit processes will influence the quality of the water more directly. Some examples of important control actions that are of great importance for the quality of the operation of the plant, and thus the effluent water quality are provided. The interested reader may refer to [7] for a detailed description of the biological processes in wastewater treatment.

The Aeration Process

In the biological reactor, *organic matter* is removed by biological oxidation. Microorganisms feed on the carbon and the oxygen in the water. Some of the organic matter is used to maintain the organisms, while the rest is converted into carbon dioxide. The process is very similar to the way in which humans use food (organic matter) to generate energy. The energy is released when the organic matter is converted into carbon dioxide during respiration. Approximately half of the organic matter is used to increase the body mass of the microorganisms, and the rest is converted into carbon dioxide.

Most of the *nitrogen* that arrives at the wastewater treatment plant comes as ammonium (60–80%).

The remainder is predominantly bound to both soluble and particulate organic matter. The nitrogen bound to organic matter is released as ammonium. The particulate material is hydrolyzed and produces soluble organic matter and nitrogen.

The removal of ammonium takes place in two principal stages: nitrification and denitrification. The nitrification process oxidizes ammonium into nitrate, while denitrification reduces nitrate into free nitrogen. The nitrification process is performed by a group of microorganisms called nitrifiers. Nitrifiers are autotrophs, that is, they can use carbon dioxide (which the water contains) to grow, as opposed to the heterotrophic microorganisms, which feed on easily degradable organic matter. Oxygen is a prerequisite for this process. Conditions in which dissolved oxygen is present are called “aerobic conditions.”

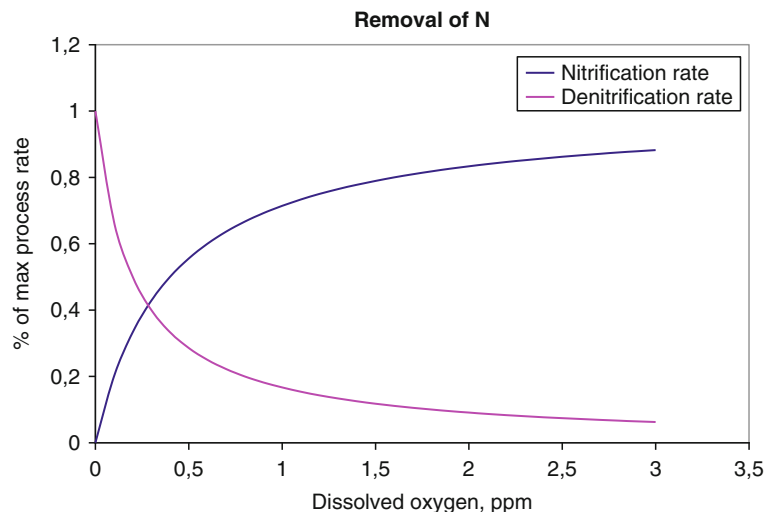
Dissolved Oxygen Control

It is obvious that the concentration of dissolved oxygen (DO) is crucial for the oxidation of both organic matter and of ammonium. Both these processes take place in

the same reactor. If the DO concentration in the aerator is too low, then the growth rate of the organisms will be limited, as illustrated in Fig. 5. If the DO level is very high then the increase of the growth rate is only marginal. Therefore, a too high DO level will make the operating costs too high.

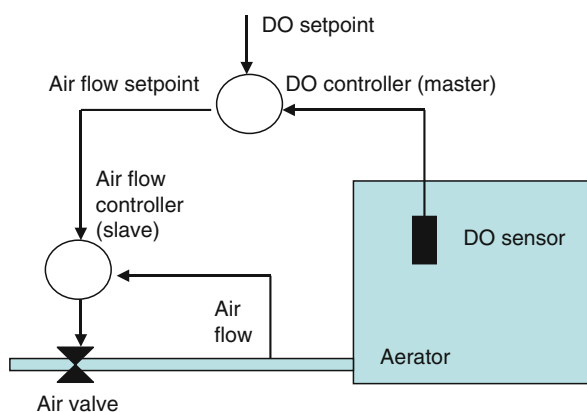
To begin with, it is assumed that the desired DO concentration is constant. The microorganism growth rates will increase for an increasing organic or ammonium load. This can be noticed as an increasing oxygen uptake rate (OUR). In other words, if the load increases, then the DO concentration will decrease due to the increase of the oxygen uptake rate.

The control of aeration has been the subject to considerable research since 1970s, when the dissolved oxygen (DO) sensors reached a level of robustness and precision suitable for feedback control [2]. Today, the control of DO to a setpoint is considered a mature technology from the methodological point of view. Remember, that the actuators are crucial. Thus, the compressor has to allow a variable airflow that is required by the control system. Still there may be



Water and Wastewater Operation: Instrumentation, Monitoring, Control and Automation. Figure 5

The principal behavior of the growth rate of microorganisms as a function of the DO concentration. For aerobic microorganisms the growth rate will increase with the DO concentration, while for anoxic organisms the growth rate is highest for an oxygen-free environment



Water and Wastewater Operation: Instrumentation, Monitoring, Control and Automation. Figure 6
Structure of a standard dissolved oxygen (DO) cascade control loop

inadequate capacity of the blowers for very high loads. From a control point of view, this is a common problem: The control variable most often has a lower and upper bound.

Figure 6 illustrates the structure of a typical DO control system.

A DO controller is based on the measurement of the DO concentration in the reactor. It is, for simplicity, assumed that the reactor is a completely mixed aerator. Then the DO sensor will represent the concentration in the whole aerator. A too low DO concentration has to be compensated by an increase in the airflow rate. Consequently, a too high DO concentration will be decreased by decreasing the airflow rate. In the control system, the DO concentration is compared with the DO setpoint and the DO controller (the master) will calculate the necessary airflow change required to change the DO concentration toward the desired value. However, the DO controller does not directly manipulate the air valve. Instead, the desired airflow is given as a setpoint to a second controller, the airflow controller (the slave). This controller receives the airflow rate measurement and compares it with the desired airflow. This difference will then make the actuator (a compressor or a valve) change the airflow to the desired value. The loop is called a *cascaded control loop* and is the standard configuration in most DO control systems.

Dissolved Oxygen Control Based on Ammonia Measurements

The oxidation of ammonia via the nitrifiers is a slower process than the oxidation of the organic matter. This means that the organic matter will disappear faster than the ammonia. Therefore, the decrease of the ammonia concentration is of primary interest. With the development of nutrient sensors it has been possible to extend the dissolved oxygen control to allow for an online adjustment of the level of oxygen supply. It means that the appropriate DO setpoint can be determined by online ammonia measurements. For details, [1], [2], [3] should be referred to.

If the aerator is designed like a plug flow reactor then the ammonia concentration is high at the inlet and will gradually decrease toward the outlet. If the ammonia concentration approaches zero well before the outlet, then the operation has been too ambitious and the last part of the reactor is not needed for the oxidation process. This behavior can be corrected with DO control. If the DO setpoint is lowered, then the growth rate of the nitrifiers will be slightly smaller. Thus, the DO setpoint should be decreased so much that the ammonia will approach zero close to the inlet. As a consequence, the energy consumption will decrease.

Conversely, if the ammonia does not approach zero at the outlet there may be several reasons. The first one to be tested is if the DO setpoint has been too low. Then the setpoint can be increased (but only to a maximum value) to ensure that the growth rate of the organisms has been sufficient. If the ammonia still does not approach zero, then the aerator is overloaded at that time and the nitrification capacity is insufficient.

A DO setpoint controller can be a simple PI controller. The setpoint of the controller is the desired ammonia value at the outlet of the aerator. The controller produces a control signal, which will serve as the desired DO setpoint. This signal is then sent to the DO control system (Fig. 6). Now the DO setpoint becomes a time-varying value and no longer a fixed value given by the operator. The resulting controller has become a hierarchical

structure of three controllers working in cascade by the master–slave principle.

This kind of ammonia controller DO setpoint has been introduced in many plants. The energy savings can be significant, especially during low load periods. In the experiments reported in [8], the savings were 28% compared to the parallel line where a constant DO setpoint was applied. This corresponds to a significant part of the operating costs and can motivate the extra cost of an ammonia analyzer.

Phase Length Control in Sequential Batch Reactors

Sequential batch reactors (SBRs) are quite popular in many places and they offer interesting control opportunities. During the biological oxidation phase, there is no influent to the reactor. As a result the organic matter content will decrease with time. Likewise the ammonia concentration will decrease as a function of time, but slower than the organic matter. The ammonia concentration as a function of time looks like the ammonia concentration as function of the reactor length in a plug flow reactor. Ammonia is oxidized to nitrite and then to nitrate, so the decrease in ammonia corresponds to an increase in nitrate concentration.

A great advantage with an SBR compared to a continuous flow reactor is that the ammonia concentration change can be measured online with just one sensor. The oxidation process is completed when the ammonia concentration approaches zero.

From a control point of view, there is an even more attractive way to determine the phase length of the oxidation in the reactor. A DO control system will serve the same purpose in an SBR as in a continuous flow reactor. Now it is assumed that the DO is to be kept constant during the oxidation phase. In the beginning, the oxygen uptake rate is high. Consequently, the required airflow rate is high. As the reaction goes toward completion the required airflow rate will decrease gradually and when the ammonia has been oxidized then the air flow demand will no longer

decrease but will approach a constant low value. This behavior can determine the end of the oxidation phase.

If the phase length of an SBR is controlled by a timer, as in open loop control, then the completion of the reaction cannot be guaranteed. As a result, many SBR operations are made with a safety margin. This means that the ammonia level approaches zero well before the end of the oxidation phase. This of course will cost money, both in terms of energy and – more important – in terms of lower plant capacity.

Control of Denitrification

In the denitrification process nitrate is converted into free nitrogen gas. The denitrification process is a reduction process and is performed by heterotrophic microorganisms, denitrifiers. They are the same type of organisms that use oxygen to oxidize organic matter when oxygen is present. Conditions in which oxygen is not present are called anoxic. Under such conditions the heterotrophic organisms can convert organic matter and nitrate into free nitrogen, carbon dioxide, and increased body mass.

Nitrification and denitrification cannot take place at the same time, since the former is aerobic and the latter is anoxic. Treatment plants are therefore designed to allow these processes to occur either in different places (as in a continuous flow system) or at different times (as in a sequential batch reactor).

The concentration of dissolved oxygen also governs the denitrification process, and the denitrification rate will be hindered by any oxygen present, as illustrated in Fig. 5. Therefore, it is extremely important to prevent oxygen from entering the anoxic process. Oxygen can accidentally enter such an environment via the surface of water, via recirculated water from the aerobic process tanks or from filter backwashing.

The denitrification requires not only nitrate but also easily degradable organic matter. Therefore, it must take place where the organic matter is

present. Since the inlet water contains a high concentration of organic matter the denitrification can paradoxically be placed before the nitrification. From the outlet of the nitrification process, nitrate-rich water is recirculated into the anoxic reactor.

First, consider a plug flow like anoxic process with continuous flow. The concentration of nitrate will decrease along the reactor and should approach zero toward the end of the anoxic reactor. The outlet concentration can be measured with a nitrate sensor. As in the aerobic system, if the concentration approaches zero well before the outlet, then there is an overcapacity. In this case, the control action is different than for the aerator.

If the nitrate concentration in the anoxic reactor has approached zero before the outlet, then the recirculation flow rate of nitrate-rich water can be increased until the nitrate concentration approaches zero close to the anoxic reactor outlet. In this way, the capacity of the anoxic reactor can be increased. Similarly, if the nitrate concentration is too high at the outlet of the anoxic zone then too much nitrate-rich water has been recirculated and the recirculation flow rate must be decreased. More details are found in [8].

Nitrate sensors are commercially available today, so a control system using this information can be readily installed. The setpoint of the controller is simply the desired nitrate concentration value close to the anoxic reactor outlet. The control signal is connected to the speed controller of the recirculation pump.

In a sequential batch reactor, the phase length of the denitrification can be readily controlled. A single nitrate sensor can measure the nitrate concentration. When it approaches zero the anoxic phase is completed. More results on phase length control in SBR systems are found in [9].

The redox potential (oxidation-reduction potential, ORP) contains certain distinctive features that can be used for denitrification control. Denitrifying activity has been found to decrease with an increased redox potential that corresponds to an increased oxygen concentration. The relationship between

denitrifying activity and the redox potential has been found to be more or less linear. However, during transient conditions the relationship may be far from linear. There can be a distinct change in the slope of the redox when the nitrate has been consumed, but there is much debate about the efficiency of this measurement, especially under transient conditions. However, a lot of results have been reported, see [10].

Monitoring

To track the current process operational state via the instrumentation is called *monitoring*. Monitoring has to start with the validation of measurement data. Even reliable instrumentation can fail during operation, which can have serious consequences if the instrumentation is used in closed loop control. There are many data screening methods that should be applied, see [1, 2]. Missing data have to be detected, limits of the signal amplitude and its rate of change should be tested, and outliers should be identified. If a signal is expected to be constant then the variability of the signal can give information. Deviations of the signals larger than $\pm 2\sigma$ and $\pm 3\sigma$ from the mean should be noted. It is obvious that deviations large than 3σ ought to be observed carefully and suitable operations have to be implemented.

In some cases, it is possible to perform cross-validation on measurements from more instruments if any correlation is expected, [1, 11]. If the confidence in a measurement decreases, it may be possible (on a short-term basis) to use an estimated value, but eventually control must be set to a default scheme until confidence in the measurement has been restored.

In a sophisticated treatment plant there is a huge data flow from the process. More instrumentation and new instrumentation development will further provide more data. Unlike humans, computers are infinitely attentive and can detect abnormal patterns in plant data. The capability of computers to extract patterns (useful information) is rarely utilized beyond simple graphing. Information technology is not commonly

used to encapsulate process knowledge, that is, knowledge about how the process works and how to best operate it. Process knowledge is typically built up from the experience of operators and engineers but all too often disappears with them when they leave. If process knowledge can be encapsulated, then not only is it retained but the computer can also assist decision-making in plant operation. The potential of substantial operator support for diagnosis and for corrective actions is there and has been demonstrated, but it needs to be adopted by the water and wastewater industry.

The early detection and isolation of faults in the biological process are very effective because they allow corrective action to be taken well before the situation becomes unfavorable. Some changes are not very obvious and may gradually grow until they become a serious operational problem.

Most signals are corrupted by noise, coming either from the instrument itself or from the process. Such noise should be eliminated and signal processing offers many methods, both simple and sophisticated. A simple moving average is often used, but more sophisticated methods can offer much better possibilities to extract information from a noisy signal.

There are many variables that are not directly measured but still of great interest. Here two examples, oxygen uptake rate and nitrification rate, are discussed.

The oxygen uptake rate (OUR) is a function of the biological oxidation of both organic matter and of ammonia nitrogen. It is obvious that the OUR has a major influence on the DO mass balance in the water, and an increase of the OUR must be compensated by an increase in the airflow rate. From a simple mass balance of DO the OUR can be calculated from the airflow rate, the oxygen transfer rate (from gaseous to dissolved oxygen), and from the DO concentration. This makes it possible to track the activity of the organisms in the biological reactor [2].

The reaction rate can be readily calculated in an SBR. Having just one ammonia analyzer makes it possible to track the ammonia concentration as function of time. From this information, it is possible to calculate the rate of change of ammonium. This in turn

can inform the operator about the performance of the nitrification. Full-scale experiences are reported in [12].

Plant Wide Control

There are several driving forces that motivate the challenge to integrate all wastewater operations in a drainage area. From a European perspective the need for integrated modeling and control has increased dramatically by the EU-Water Framework Directive, which forms a basis for water policy for the EU and associated countries. It focuses strongly on ecological criteria and strategic integrated management. The system boundaries are extended to include river management at a basin scale as well as point-source and diffuse-source pollutions. A new set of integrated and harmonized tools supporting the intention of the WFD is required.

Instrumentation, control and automation (ICA) play an important role for the integration and concepts like *integrated* control, *system-wide* control, and *plant-wide* control are often mentioned. Both the sewer system and the wastewater treatment plant (WWTP) have an impact on the receiving water. Combined sewer overflows (CSOs) as well as the wastewater treatment effluent will influence the quality of the receiving water, the former at discontinuous events and the latter on a continuous basis. Here the concept of plant-wide control is defined as the combined sewer and WWTP control and operation. The objective is to make maximum use of the system capacity in order to meet large disturbances, where the ultimate purpose is to minimize the combined load to the receiving water.

For the sewer the goal is to minimize the amount of CSO. The aim of the control of the WWTP is to satisfy the effluent requirements while minimizing the operational costs [2]. During storm conditions, these goals may be difficult to reach. It is obvious that the control of the sewer system – isolated from the WWTP operation – will lead to suboptimal solutions. If the goal of the sewer operation is to minimize the CSO, the WWTP may soon be overloaded. The result of such an overload is firstly a short-term impact due to the increased effluent

pollution. Then a possible washout of organisms may cause a long-term impact by making the WWTP operation less efficient during an extended period of time.

In order to implement plant wide control, the interactions between the system parts and their dynamic characteristics have to be understood. This includes choosing the best locations of sensors and actuators, which is the basis for the control structure. ICA has a large impact on integrated operation and compromises have to be made when integration of different unit operations is getting realized. Modern plant computer systems allow a wide distribution of control but this is not always met by corresponding control structures.

Integration aims at minimizing the impact on the receiving water, while ensuring a better resource utilization. The system resilience is an important factor. This includes its ability to attenuate disturbances, but should also reflect its sensitivity to major disturbances or even purposeful and harmful attacks. In the integrated approach, the ultimate goal is to formulate some criterion for the receiving water and its ecological quality while satisfying various economic and technical constraints. There is a great challenge to relate this performance to the plant effluent and possible sewer overflow. Performance measures of the plant operation that relate effluent quality to the resources that are needed to obtain it, such as energy, chemicals, and other material and operating costs, are needed. This is not yet solved satisfactorily, but promising research is in progress, such as [13]. Models are getting developed to find strategies to dynamically find maximum WWTP loading according to continuous monitoring and prediction of the operational state. Another aspect is storage management (in the sewer system and in retention tanks), not only during storms but also during normal operations. By mixing different types of wastewater to compensate, for example, for nutrient deficit or overload the capacity of the plant can be maximized.

All integration means some kind of compromise. This is apparent and integrations at various levels have been applied for a long time. If there were no interactions, then the individual optimization of each subprocess would be the best strategy. Having couplings in reality means that an integrated approach is

better than if each one of the processes were controlled separately. Another way to phrase the compromise is to call it a multi-criterion index: Various performances have to be weighted and compared with each other. This is a common approach in integrated water management [14]. The idea is illustrated with some examples:

- The interaction between the aerator and the settler is a classical integration problem, reflected in the compromise that has to be done in return sludge flow rate control.
- The anoxic zone in a pre-denitrifying plant interacts closely with the nitrifying aerator. Oxygen-rich water is recirculated from the aerator to the anoxic zone. The DO level has to be a compromise between sufficiently good nitrification and denitrification. This has been discussed in [Control of Biological Oxidation](#) and [Control of Denitrification](#).
- Recycle streams interconnect various parts of a treatment plant. As noted in [Incentives for Control](#), supernatants from the sludge treatment are most often highly concentrated in nutrients and have to be synchronized in time with the plant influent load.
- The target for the sludge production is not the same in different plants. Sometimes the target is to maximize the biogas production, at other plants the sludge production needs to be minimized.
- In the combined sewer and WWTP operation, the individual system operations are sometimes in conflict, so the overall goal of minimizing the load to the receiving water has to overrule the individual goals [14].

Future Directions

ICA will play an increasing role in wastewater treatment systems. It has been noted that the introduction of ICA can increase the capacity of a treatment plant, both a continuous flow system and a sequential batch reactor system. To meet future environmental and efficiency demands it is needed to increase the integration of plant design

and operation. Today, most often the ICA is added to an already existing design, which often leads to insufficient performance.

The instrumentation development will favor more implementations of control systems and lead to more consistent and economical operation of the treatment plants. Furthermore, the actuator equipment has to get more attention. Variable speed control is a proven technology today, but often not observed by the designers. Furthermore, efficient design of pumps and compressors is a prerequisite for economic operation of the plants.

Today most design is based on safety factors for the operations and the disturbances to the plant. There is a clear ambition to gradually make the design more model based, where various risk factors can be taken into consideration. This will result in a better coupling between design and operation.

Plant wide control still needs to be developed. More stringent energy consumption requirements will be one driving force for this development. In [Plant Wide Control](#), the compromise of sludge production was mentioned. A wastewater treatment plant not only consumes electrical energy. It can also produce energy as biogas in the anaerobic digestion. This will immediately require a compromise between the biological oxidation and the anaerobic digestion. The organic matter in the influent is needed to some extent for the nitrification and the denitrification. However, in order to maximize the biogas production as much organic matter as possible should bypass the aerobic oxidation and be forwarded to the digester. Thus, a plant size production planning will be required.

Today many of the unit processes in the wastewater treatment plant can be controlled, but methods and criteria need to be further developed to achieve the integration of unit processes within the plant and to further integrate the plant into an urban water treatment system.

Bibliography

1. Olsson G, Nielsen MK, Yuan Z, Lynggaard-Jensen A, Steyer JP (2005) Instrumentation, control and automation in wastewater treatment systems. Scientific and technical report No.15, IWA Publishing, London, UK, ISBN 1-900222-83-3
2. Olsson G, Newell B (1999) Wastewater treatment systems. Modelling, diagnosis and control. IWA Publishing, London, ISBN 1-900222-15-9
3. Olsson G (2008) Process control. In: Henze M, van Loosdrecht M, Ekama GA, Brdjanovic D (eds) Biological wastewater treatment – principles, modelling and design. UNESCO, IWA Publishing, London. ISBN 9781843391883
4. Vanrolleghem PA, Lee DS (2003) On-line monitoring equipment for wastewater treatment processes: state of the art. *Water Sci Technol* 47(2):1–34
5. Marlin TE (2000) Process control. Designing processes and control systems for dynamic performance, 2nd edn. McGraw-Hill, New York. ISBN 0-07-039362-1
6. Wilkie J, Johnson M, Katebi R (2002) Control engineering. An introductory course. Palgrave, New York. ISBN 0-333-77129-X
7. Henze M, van Loosdrecht M, Ekama G, Brdjanovic D (eds) (2008) Biological wastewater treatment – principles, modelling and design. UNESCO, IWA Publishing, London, ISBN 1843391880
8. Ingildsen P (2002) Realising full-scale control in wastewater treatment systems using in situ nutrient sensors. Ph.D. thesis, Dept of Industrial Automation, Lund University, Lund, Sweden (access via www.iea.lth.se)
9. Rönner-Holm SGE, Holm NC (2009) Special automation and regulation strategies for enhancing sequencing batch reactor (SBR) performances. *Water Sci Technol* 60(5):1161–1172
10. Cecil D (2008) The control of denitrification time in full scale by the automatic detection of the low nitrate bend in the redox curve. *Water Sci Technol* 57(7):1095–1101
11. Lynggaard-Jensen A, Frey M (2002) Use and reporting of data – data handling and validation. In On-line monitoring for drinking water utilities. American Water Works Association Research Foundation, Denver, Colorado (2002). ISBN 1-58321-183-7
12. Rosen C, Ingildsen P, Guildal T, Nielsen MK, Jacobsen BN, Ønnerth T (2004) On-line estimation of nitrification and denitrification capacity at Avedøre wastewater treatment plant. In: Proceedings of the international conference upgrading of wastewater treatment plants (AquaTech2004). Amsterdam, September 30–October 1 2004
13. Benedetti L, Prat P, Nopens I, Poch M, Turon C, De Baets B, Comas J (2009) A new rule generation method to develop a decision support system for integrated management at river basin scale. *Water Sci Technol* 60(8):2035–2040
14. Vanrolleghem P, Fronteau C, Bauwens W (1996) Evaluation of design and operation of the sewage transport and treatment system by an EQO/EQS based analysis of the receiving water imission characteristics. In: Proceedings of WEF conference urban wet weather pollution. Québec, 16–19 June 1996, pp 14.35–14.46

Water Reclamation System and Micropollutants

JIANG YONG HU, SE-KEUN PARK
Division of Environmental Science and Engineering,
National University of Singapore, Singapore

Article Outline

Glossary
Definition of the Subject
Introduction
Endocrine Disrupting Compounds
Removal of Endocrine Disrupting Compounds in
Membrane Bioreactor Systems
Future Directions
Bibliography

Glossary

Activated sludge process A biological method of wastewater treatment that is performed by a variable and mixed community of microorganisms in an aerobic aquatic environment.

Biodegradation The reaction processes mediated by microbial activity.

Endocrine disrupting compounds (EDCs) An exogenous agent that interferes with the synthesis, secretion, transport, binding, action, or elimination of natural hormones in the body, which are responsible for the maintenance of homeostasis, reproduction, development, and/or behavior.

Hydrophobicity The physical property of a molecule that is repelled from a mass of water.

Membrane bioreactor (MBR) The combination of a membrane process like microfiltration or ultrafiltration with a suspended growth bioreactor.

Micropollutant Pollutant, which exists in very small traces in water.

Sludge retention time (SRT) An operational parameter, which measures how long the sludge can remain in a reactor.

Sorption The action of absorption or adsorption.

Water reclamation The various treatment systems and their process trains, which further treat the used water for different reuse purposes.

Definition of the Subject

Endocrine disrupting compounds (EDCs), sometimes known as “hormonally active agents” or “endocrine modulators,” are micropollutants with the potential to elicit negative effects on the endocrine systems of humans and wildlife and are emerging as a major concern for water quality. In the past few decades, more and more chemicals were monitored as the source of estrogenic activity in wastewater and the primary sources of estrogenic activity in wastewater were found to be natural estrogens such as estrone, 17 β -estradiol and estriol, and synthetic compounds like 17 α -ethinylestradiol, alkylphenols, and alkylphenol ethoxylates. EDCs are either partially retained in the sludge, or metabolized to a more hydrophilic but still persistent form, and discharged consistently from the wastewater treatment plant into the receiving waters, thereby causing endocrine disrupting effects to aquatic organisms even at very low concentrations (ng/L). As conventional processes are not able to fully and systematically eliminate EDCs to a satisfactory level, membrane bioreactor (MBR) is considered as a promising technology in water reclamation to achieve further removal of EDCs due to specific design characteristics (such as high sludge age, high biomass concentration, and complete particle and microorganism retention). The MBR, which integrates biological process with membrane filtration, is an effective water reclamation process developed for recent decades and surpasses the limitations of the conventional activated sludge process. Therefore, with respect to the wastewater reclamation and reuse, elimination of EDCs in wastewater is typical state-of-the-art application of the MBR technology.

Introduction

There is a growing concern worldwide for the potential effects of endocrine disrupting compounds (EDCs). These EDCs can mimic or modulate the action of a particular natural hormone in human and wildlife. They can also attach onto receptors intended for a particular hormone, interfering and altering the normal binding of the hormone to its receptors.

Many adverse health and environmental problems are found to be linked to EDCs. Colborn et al. [1] observed various disturbing consequences relating to

the alteration and interference of endocrine functions in humans and wildlife. These effects of EDCs depend on several factors such as its type, dosage, and exposure duration. The adverse effects caused by EDCs have raised concerns pertaining to the levels of EDCs present in the environment. To facilitate the excretion of urine, the female body primarily excretes estrogens in a biologically inactive form as sulfate and glucuronide conjugates, which may exist in the influents and effluents of sewage treatment plants (STPs) [2]. Others have also reported that estrogenic activities are present in sewage and treated effluents [3, 4]. These treated effluents are often discharged into surface water bodies such as rivers, seas, or lakes, to which humans and wildlife are often exposed.

EDC concentrations in treated effluents are sufficient to induce vitellogenin biosynthesis in male fish [5], alterations of endocrine reproductive systems, decreased fertility and growth, poor hatching, egg shell thinning and abnormal thyroid function for birds, reptiles, and mammals [6, 7], and malformations of newborns, undescended testicles, abnormal sperm, low sperm counts, abnormal thyroid function, breast cancer, testicular prostate cancer for humans [8–10]. A classic example is the feminization of male fish like trout downstream of sewage treatment plant in some areas of England and the USA [5, 11].

The EDCs present in the sewage can be categorized into two main groups, namely, estrogens and alkylphenols. The predominant estrogens can be subdivided into natural estrogens and synthetic estrogens. Natural estrogens such as estrone (E1), 17 β -estradiol (E2), and estriol (E3) are present in the human excreta whereas synthetic estrogens such as 17 α -ethinylestradiol (EE2) can be found in contraceptives. Alkylphenols like bisphenol A (BPA) and nonylphenol (NP) are released from manufacturing industries. Natural estrogens, i.e., E2 and E1, are effective at the nanogram per liter level [12, 13], while most other EDCs are biologically active at the microgram per liter level. For example, levels reported to affect fish were in the range of 1–4 ng/L for EE2 [14, 15], 1–10 ng/L for E2, and 1–10 μ g/L for alkylphenol in vitro [16], and in vivo [13].

To investigate whether the effluent of most STPs contains EDCs and whether these EDCs lead to the above effects, several investigators have attempted to

quantify the concentrations of those compounds in the effluent of municipal or industrial STPs. The steroid estrogens reportedly ranged from 3 to 9 ng/L for E1 [4, 17], 0.1 to 5 ng/L for E2 [18, 19], and 0.1 to 9 ng/L for EE2 [4, 17, 20], and 4-NP and BPA were detected in the effluent at levels from 0.16 to 0.36 μ g/L [21].

It is clear that STPs are not able to treat the wastewater to a satisfactory level. As a result, methods to improve their removal efficiency should be explored. Membrane bioreactor (MBR) is an attractive option mainly due to its improved effluent quality. Holbrook et al. [22] studied a membrane bioreactor pilot plant and a STP receiving the same domestic wastewater. The influent 17 β -estradiol equivalent concentration (EEQ) was 16.25 ng/L, while the effluent EEQ was 4.5 ng/L for MBR and 7 ng/L for STP, on average. It showed that the MBR achieved a higher reduction in EEQ concentrations than the conventional treatment system.

MBR system is a biological treatment process that combines with membrane separation process in order to biologically treat wastewater and physically separate solid and liquid in an integrated step. The integration of biological water treatment with its filtration across porous membranes, driven by a pressure difference, was first applied for the separation of activated sludge by a membrane and its recycling to the aeration tank [23]. Its technological and economical efficiency (including the decrease in the membrane cost) have improved steadily since, allowing for more practical applications. Subsequently, MBR has been extensively adopted from the mid-1990s for the treatment and reclamation of municipal and industrial wastewaters.

Application of membrane separation techniques for biosolid separation can overcome the disadvantages of the sedimentation tank and biological treatment steps. The membrane offers a complete barrier to suspended solids and yields higher quality effluent. The outstanding merits of MBR over the conventional activated sludge systems include excellent quality of treated water, small footprints, complete solid–liquid separation, high biomass concentration, prolonged sludge ages, less sludge production, and better process reliability in the bioreactor [24, 25]. The main advantage of MBR system is a complete retention of the biomass within the bioreactor as separation is achieved via membrane filtration. Thus, MBR can be operated

at much higher biomass concentrations providing a higher biological activity per volume. A very long sludge retention time (SRT) can also be maintained, allowing for the establishment of a more diverse community with broader physiological capabilities. This feature is of particular importance for EDCs, which usually need high biomass and long periods for their complete biological degradation/transformation to harmless products.

It is generally agreed that MBR performs better than conventional STP for biological removal of bulk organic matter. Yet, the fate of EDCs during MBR treatment and the mechanisms relevant for their removal are much less documented. This article reviews the removal of EDCs during wastewater reclamation applying MBR and STP, as well. Biodegradation of EDCs is discussed as a major mechanism, as well as their removal by sorption onto sludge.

Endocrine Disrupting Compounds

General Characteristics of EDCs

Natural estrogens E1, E2, and E3 have an aromatic A ring, basic C18 skeleton estrane and no carbon side-chain at C17 (Table 1). These are composed of the four rings: a phenol, two cyclohexanes, and a cyclopentane. The difference in the compounds within the C18 group lies in the configuration of the cyclopentane ring (ring D) at positions C16 and C17. For example, E1 has a carbonyl group on C17, E2 has a hydroxyl group on C17, and E3 has two hydroxyl groups on C16 and C17. Hydroxyl and carbonyl functional groups make them capable of participating in hydrogen bonding as a proton donor or proton acceptor species. Table 1 shows the physical and chemical properties and the biological potency of estrogens. From the octanol–water partition coefficient ($\log K_{OW}$) values, E1 and E2 are considered to be more hydrophobic than E3. The estrogenic activity of E2 is much stronger than E1 and E3 as revealed by the results of yeast assay [27] and feminization of male fishes [28].

The synthetic estrogen EE2, which is used for estrogenic deficiencies treatment and oral contraceptive pill, has lower solubility and is somewhat more hydrophobic than the other steroid estrogens ($\log K_{OW}$ values shown in Table 1). The major difference between EE2

and the natural estrogens is the long persistence in the environment with a dissipation time in the order of 20–40 days in river systems [29]. Also, EE2 is known as an extremely potent estrogen, and its strong estrogenicity was shown in both in vitro and in vivo studies [15].

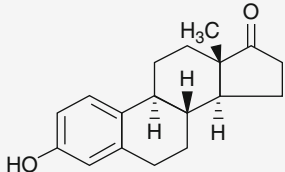
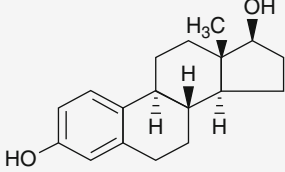
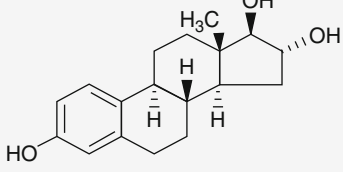
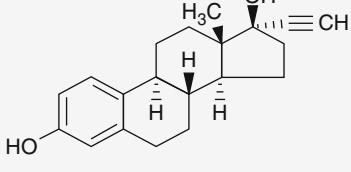
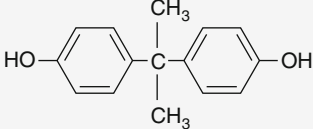
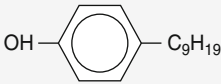
Estrogens are usually released from conjugated forms with glucuronide (G) and/or sulfate (S) groups linked to the H in the –OH group of the steroid ring. It is known that most of estrogen conjugates exist in influent and effluent of STP with high frequency for both glucuronide and sulfate estrogens. However, the conjugated estrogens have low estrogenic potency compared with the parent compounds [30].

BPA and NP are widely used in a variety of industrial and residential applications and stimulate production of the biomarker of estrogenic activity, vitellogenin, in male fish [31–33]. BPA and NP usually possess similar endocrine disrupting activities due to phenyl ring in their chemical structures.

BPA is a monomer utilized to manufacture polycarbonate plastic and epoxy resins. It is a moderately water-soluble compound (300 mg/L at 20°C) and dissociates in an alkaline environment (pK_a 9.9–11.3). Numerous toxicological and biochemical studies have confirmed that BPA has estrogenic properties and an agonistic effect toward the estrogenic receptor. Estrogenicity of BPA was first reported in 1936 by Dodds and Lawson and subsequently identified as a result of studies on the release of estrogenic material from autoclaved polycarbonate flasks [34]. The estrogenic activity of BPA is less potent than that of E2 in in vitro assays [34–36]. However, the results of in vivo studies on the estrogenicity of BPA are conflicting in terms of the effective dose.

NP is a chemical substance with a phenolic ring and an attached lipophilic linear nonyl chain or, more usually, branched nonyl group. 4-NP, in which the side chain is attached to the carbon directly opposite the hydroxyl group, is the most common member of NP family. NP is a hydrophobic compound with a $\log K_{OW}$ of 4.5 and low solubility in water. This exhibits minor estrogenic potency. In stimulating estrogenic activity, NP was 1,000–100,000 times less potent than E2. In vitro studies showed that NP could activate the estrogen receptor with a potency of 5,000–7,000 times lower than that of E2 [37].

Water Reclamation System and Micropollutants. Table 1 Physicochemical properties and estrogenic activity of estrogens

EDCs	Chemical structure	Molecular weight (g/mol)	Water solubility at 20°C (mg/L)	Log K_{ow}	E2 relative potency (YES)	E2 relative potency (E-screen)
Estrone (E1)		270.4	13	3.43	0.38	0.01–0.096
17 β -Estradiol (E2)		272.4	13	3.94	1.0	1.0
Estriol (E3)		288.4	13	2.81	0.024	0.1–0.25
17 α -Ethinylestradiol (EE2)		296.4	4.8	4.15	1.19	0.91–1.0
Bisphenol A (BPA)		228.0	300	3.40	1.0×10^{-5}	1.26×10^{-5}
Nonylphenol (NP)		220.0	5.43	4.48	1.7×10^{-4}	7.8×10^{-5}

Source: Teske and Arnold [26]

Estrogenic potencies from yeast estrogen screen (YES) bioassay data and/or the E-Screen bioassay

Occurrence of EDCs and Their Conjugates in Aquatic Environment

The presence of EDCs in the environment has become a concern because they may interfere with the reproduction of man, livestock, and wildlife. So far, a number of studies have reported the occurrence of EDCs in measurable concentrations (ng/L or µg/L) in wastewater, surface water, groundwater, and even drinking water. Various reported concentrations for major EDCs in these different aquatic environments are listed in Table 2. The highest concentrations of EDCs have been observed in wastewaters. Some studies showed a very wide distribution of EDCs with some sporadic occurrence [17, 44, 45, 54, 65]. The levels of natural EDCs appear to be much higher than those of synthetic estrogen EE2. In the studies, E1 was detected frequently with high concentrations in effluent of STPs [17, 55, 68, 69]. In some cases, EE2 has been reported to more frequently occur in the effluents of STPs with concentrations ranging from <5 to 178 ng/L [54]. Obviously, alkylphenol chemicals are present at much greater concentrations than estrogens in aqueous environmental samples. Some research exceptionally showed that natural and synthetic estrogens as well as alkylphenols can be detected in drinking water samples at pg/L or ng/L levels [40].

From Table 2, it can be seen that the EDCs pass through STPs without being completely removed and remain with fluctuating concentrations in effluent. These discharges from STPs may be the main reason for the wide distribution and occurrence of EDCs in surface waters, ground waters, and even in drinking waters. In particular, concentrations of some EDCs such as E1 are higher in effluent than in influent [17, 68, 69], probably due to microorganism-induced transformation (e.g., from E2 to E1). Incomplete removal and degradation of EDCs by existing STPs not only results from the fluctuation of EDC levels in the influent, but the processes of the STPs, and as well as operational conditions [70].

A number of past studies have shown the estrogenic activity in STP effluents using fish or other biological assay systems [12, 71]. In the study carried out by Purdom et al. [12], EE2 could induce vitellogenin (VTG) in male rainbow trout at concentration as low as 0.1 ng/L. In addition, Routledge et al. [13] concluded

from the dose–response curves in male rainbow trout that the threshold response for VTG production probably occurred between 1 and 10 ng/L for E2, and between 25 and 50 ng/L for E1, respectively. Indeed, alkylphenols can have estrogenic effects in fish at concentrations from 1 to 10 µg/L [13, 14]. Jobling and colleagues [14] showed that these chemicals were also estrogenic in fish, and therefore might be partially responsible for the estrogenic effects (e.g., intersexuality) observed in wild fish. Recent report on VTG induction by 4-NP demonstrates the presence, in different fish species, of a lowest observed effect concentration identifiable by the range of 10–20 µg/L NP [72]. From these findings, it is emphasized that EDCs that are present in sufficient quantity in treated as well as raw wastewater may be discharged with effluents into receiving waters to produce estrogenic effects.

Treatment of EDCs in Conventional Sewage Treatment Plants

Since EDCs have been identified and quantified in a wide variety of environments associated with municipal or industrial STPs effluents, the removal of these substances in these treatment plants is becoming an increasingly important environmental issue worldwide. The activated sludge process is commonly used to treat wastewater in the world. There are many research data suggesting that the activated sludge STPs can remove EDCs during treatment, although its conventional purposes are the removal of bulk organic substances from wastewater. The removal performances for EDCs in activated sludge STPs are in a wide range. The removal occurs mainly through sorption by sludge and subsequent biodegradation [73]. Johnson et al. [74] reported that the activated sludge processes could remove 74% of E1 and 88% of E2. Baronti et al. [17] also reported that the removal rates of E1, E2, E3, and EE2 from wastewater in activated sludge STPs were 61%, 87%, 95%, and 85%, respectively. In the Brazilian activated sludge STPs, E1, E2, and EE2 were eliminated up to 83%, 99.9%, and 78%, respectively [75]. Moreover, efficiencies over 90% have been reported for E1, E2, and EE2 in activated sludge STPs [68, 76]. Johnson and Sumpter [73] have reviewed the literature on the fate of EDCs in activated sludge STPs. They suggested that the activated sludge

Water Reclamation System and Micropollutants. Table 2 Occurrence of EDCs in various aquatic environments at levels of ng/L

EDCs	Wastewater		Surface water	Ground water	Drinking water
	Raw	Treated			
E1	25–132 [17]	2.5–82 [17]	<0.1–3.4 [38]	ND ^a –120 [39]	0.2–0.6 [40]
	44 [41]	17 [41]	0.2–6.6 [42]	ND–4.5 [43]	
	29–670 [44]	Nd–72 [44]	<0.3–7.2 [45]	ND–1.6 [46]	
	19–78 [47]	1–96 [47]	5–12 [48]		
	20–130 [45]	<0.3–11 [45]	1.4–12.9 [49]		
	ND–18.3 [50]	ND–6.7 [50]	<17–38 [51]		
	57.8–83.3 [52]	6.3–49.1 [52]	ND–7.4 [53]		
	4–33 [54]	0–147 [54]	ND–4.6 [46]		
	230–510 [55]	ND[55]	1.1–3.0 [56]		
E2	4.0–25 [17]	0.35–3.5 [17]	<0.3–5.5 [38]	ND–45 [39]	0.2–2.1 [40]
	11 [41]	1.6 [41]	0.6–1.0 [42]	6–66 [57]	2.6 [58, 59]
	35–125 [44]	ND–30 [44]	<0.8–1.0 [45]	13–80 [60, 61]	ND–2.6 [53]
	2.4–26 [47]	0.2–14.7 [47]	2–6 [48]	ND–0.79 [46]	
	17–150 [45]	<0.8 [45]	ND–8.8 [49]		
	22.0 [62]	0.95 [62]	<0.7–1.7 [51]		
	ND–161.6 [52]	ND–5.4 [52]	1.6–15.5 [63]		
	0–11 [54]	0–158 [54]	ND–3.8 [53]		
E3	24–188 [17]	0.43–18 [17]	2–5 [48]	ND–0.16 [46]	
	72 [41]	2.3 [41]	ND–3.3 [49]		
	23–48 [48]	ND–1 [48]	ND–3.1 [64]		
	23–660 [44]	ND–275 [44]	ND–1.7 [53]		
	138–381 [65]	<1–4.9 [65]	ND–1.9 [46]		
	2–22 [54]	0–29 [54]	1.0–2.5 [56]		
EE2	0.40–13 [17]	ND–1.7 [17]	<0.1–4.3 [38]	ND–0.94 [46]	0.15–0.5 [40]
	4.9–7.1 [56]	2.7–4.5 [56]	<0.3–0.4 [45]		
	<0.3–5.9 [45]	<0.3–2.6 [45]	ND–1 [48]		
	3–70 [44]	ND–5 [44]	5.7–30.8 [63]		
	<0.7–14.4 [65]	<0.7–4.1 [65]	0.1–5.1 [40]		
	ND–1.2 [52]	ND–0.6 [52]	ND–4.6 [53]		
	0–2 [54]	0–178 [54]	1.1–2.9 [56]		

Water Reclamation System and Micropollutants. Table 2 (Continued)

EDCs	Wastewater		Surface water	Ground water	Drinking water
	Raw	Treated			
BPA	332–339 [48]	13–16 [48]	<8.8–1,000 [45]	ND–7 [48]	0.5–2.0 [40]
	250–5,620 [45]	<43–4,090 [45]	0.5–14 [40]	ND–930 [46]	5 [66]
	720–2,376 [44]	16–1,840 [44]	ND–600 [46]		
	281–3,642 [65]	6–50 [65]	15–29 [48]		
	88–438 [54]	11–1,054 [54]	65–295 [66]		
	450–3,300 [55]	610–710 [55]	ND–5.7 [67]		
NP	4,194–8,768 [48]	1,120–2,235 [48]	<110–4,100 [45]	ND–3,100,000 [39]	2.5–16 [40]
	240–19,000 [45]	550–1,500 [45]	6.7–134 [40]	ND–1,500 [46]	
	1,570–27,411 [65]	426–4,926 [65]	ND–890 [46]		
	14,728–41,207 [54]	ND–90,043 [54]	1,289–1,466 [48]		
	2,100–4,500 [55]	300–400 [55]	1.3–21.3 [67]		

^aNot detected

process can consistently remove over 85% of E2, E3, and EE2, but the removal performance for E1 appears to be less and is more variable. Slightly lower removal of E1 is probably due to the conversion E2 to E1 during the treatment and the cleavage of estrogen glucuronides and sulfates (e.g., E1-3G and E1-3S) [41, 77, 78]. In addition to estrogens, removal of BPA and NP by activated sludge STPs has been reported at different degrees in the literature [45, 48, 50, 65, 79]. In particular, a very wide range of the removal efficiency, from 10% to 99% for BPA and 33–93% for NP, is found in the references by Drewes et al. [65] and Clara et al. [79]. Their removal rates in activated sludge process were reported to be dependent on the applied sludge retention time or the mass load to plant.

It is known that natural estrogens initially excreted by humans are mainly in sulfate or glucuronide conjugates (e.g., E1-3S, E2-3G, E3-17G, etc.), and then they are changed to free estrogens. Glucuronide estrogens were reported easily changed to their free estrogens while sulfate estrogens were more recalcitrant to biotransformation [65, 77, 80]. D'Ascenzo et al. [65] reported that estrogen glucuronides spiked into wastewater from a septic tank at 25 µg/L were lost with half-lives of ~6 h and the loss of estrogen sulfates proceeded

much slower with half-lives of 2.5 days. Although estrogen conjugates are biologically less active, they are potential sources of active estrogens due to dissociation in STPs or as a result of the input of untreated wastewater directly into surface water. D'Ascenzo et al. [65] examined the fate of estrogen conjugates in activated sludge STPs and observed that E2-3G, E3-16G, and E2-3S were completely removed while a fraction of E1-3G (0.7 ng/L), E1-3S (9 ng/L), and E3-3S (2.2 ng/L) with relatively lower removal rate passed through the treatment system. Lower elimination of E1-3S in activated sludge STPs was also reported in a study by Reddy et al. [2].

In the literature on the fate of EDCs in activated sludge STPs, reported removal percentages for EDCs most often exceed 80%. However, in spite of the considerable removal rates, the low levels of EDCs that are being discharged in the final effluents of activated sludge STPs may still provoke estrogenic effects in fish and other aquatic organisms. It is generally agreed that, in most cases, the estrogens and alkylphenols will not be completely eliminated in activated sludge STPs. So, there is a need for further investigation on treatment processes to achieve EDC concentrations in effluent below estrogenic limits.

Removal of Endocrine Disrupting Compounds in Membrane Bioreactor Systems

General Aspects of MBRs

Membrane processes are regarded as key elements of advanced wastewater reclamation and reuse schemes. MBR is a promising process combination of activated sludge treatment and membrane separation. The bioreactor is operated similar to a conventional activated sludge process but without the need for secondary clarification and tertiary steps like sand filtration. MBR process is generally performed using low-pressure membranes such as microfiltration (MF) or ultrafiltration (UF) that are capable of displaying a barrier effect based on size exclusion, with pore size ranging between 50 and 10,000 nm (MF) and 1–100 nm (UF) [81]. MF and UF membranes assist activated sludge process in MBR. Suspended solids and microorganisms responsible for biodegradation are separated from the treated water by a membrane filtration. Although direct physical sieving effect of micropollutants generally of molecular weight lower than 400 g/mol is of minor importance due to the relatively large pore sizes of MF and UF membranes, hydrophobic substances tend to accumulate onto the sludge, and therefore, they are removed from the effluent.

The membranes can be placed in the external circuit of the bioreactor (known as external MBR) or they can be submerged directly into the bioreactor (known as submerged MBR). The membranes are usually of flat sheet or hollow fiber configuration if placed inside the bioreactor, or multi-tube if placed outside it. In case of the external MBR configuration, the mixed liquor is circulated to a membrane module that is outside the bioreactor, where pressure drives the separation of water from the sludge. The concentrated sludge is then recycled back into the reactor. For the submerged MBR configuration, a suction force is applied to draw the water through the membrane, while the sludge is retained on the membrane surface. Membranes act in the aeration basin as filter and improve the biological treatment of the effluent. A manifold at the base of the reactor diffuses compressed air within the reactor, providing oxygen to maintain aerobic conditions. The air bubbles also function to scour the membrane surface and clean the exterior of the membrane as they rise in the reactor. The submerged configuration is more

commonly used for wastewater reclamation than the external configuration because it is less energy consuming and less detrimental for the biomass [82].

MBR treatment of municipal wastewater yields high-quality water with reported removal percentages of 95%, 98%, and 99% (or greater) for chemical oxygen demand (COD), biochemical oxygen demand (BOD), and suspended solids (SS), respectively [83]. This is primarily ascribed to the absolute barrier provided by the membrane but can also be improved by the microbiology. Factors such as high biomass concentration and long sludge ages or SRT could be correlated with the high performance of wastewater treatment. Generally, due to the complete retention of biomass via membrane filtration, MBR is operated at much higher mixed liquor suspended solids (MLSS) concentrations, ranging from 8 to 25 g/L or even more [24, 84, 85], which provide a high biological activity per unit volume. A longer SRT can also be maintained. The sludge age in MBRs is mostly in range of 20–50 days and occasionally more than 100 days in small MBRs where almost no sludge removal may occur [86]. The prolonged SRT operation favors the growth of slowly growing specialized microorganisms, which are able to remove slowly degradable components in the reactor [87]. In a MBR, SRT can be controlled independently from hydraulic retention time (HRT), enabling MBR operation at high SRT and low HRT. The lower HRT allowable for MBR treatment may be a benefit of the higher biomass concentrations and be further supported by the fact that most of the biomass in an MBR occurs in freely suspended form [88]. This provides easier access for substrates as compared to bacteria aggregated into flocs that prevail in conventional activated sludge [89]. Due to the higher MLSS concentrations at a constant volume load, food to microorganism (F/M) ratio is usually lower in MBR. Stephenson et al. [24] reported that F/M ratio of MBR is in the range of 0.05–0.15 kg COD/kg MLSS·d. A lower F/M ratio results in substrate limitation and may force microorganisms to metabolize poorly degradable compounds [88]. This is one explanation why removal of poorly degradable pollutants may be superior in MBR systems and why this can be achieved at lower HRT. Studies on MBR note less sludge production. Bouhabila et al. [90] reported less sludge production ranging of 0.2–0.34 kg MLSS/kg COD_{removed} in MBR as compared with 0.3–0.5 kg

MLSS/kg COD_{removed} in conventional activated sludge process. It has been shown that sludge production is greatly decreased as sludge age increases [89, 91]. Lower F/M ratio and longer sludge age in the MBR are generally used to explain this lower production rate.

It is likely that biomass composition and structure in an MBR is considerably different from a conventional activated sludge. In the literature, the floc size distribution of MBR sludge and conventional activated sludge is reported to be 10–100 μm in membrane bioreactors and 100–500 μm for conventional sludge [92, 93]. Ng and Hermanowicz [94] recently showed that the MBR biomass was composed of smaller, weaker, and more uniform-sized flocs and higher fraction of dispersed microorganisms compared to the conventional activated sludge. The smaller flocs from MBRs could stimulate a higher oxygen and/or carbon substrate mass transfer and thus higher activity levels in the system. The presence of much higher amount of non-flocculating microorganisms could contribute to poorer sludge settling but better organics removal in the MBR. Zhang and coworkers [92] also found that nitrification activities in MBR processes averaged 2.28 g $\text{NH}_4\text{-N/kg MLSS}\cdot\text{h}$, which was greater than in conventional processes (0.95 g $\text{NH}_4\text{-N/kg MLSS}\cdot\text{h}$). Many studies have proven that MBR can operate as a high-rate nitrifying technology that can be applied in the nitrification of wastewater containing a high concentration of ammonia nitrogen [95].

MBR process has obvious advantages over the conventional activated sludge processes, e.g., excellent effluent quality, small footprint, less sludge production, and flexibility of operation, and becomes a promising alternative for wastewater treatment. Drawbacks of MBR operation under high MLSS concentration include limitations in oxygen diffusion leading to increased aeration cost, mass transfer limitation, and membrane fouling, which may require frequent membrane cleaning [96]. High solid concentrations may have an influence on sludge viscosity, so affecting the circulation of bulk sludge and the shear stress at the filtration cake surface [97].

MBR Performance in Removal of EDCs

It is clear that conventional processes in STPs are not able to eliminate EDCs to a satisfactory level.

As a result, methods to minimize the discharge of those compounds should be explored. MBR is an attractive option due to the potential for increased endocrine disruptor removal. Holbrook et al. [22] studied the EEQ (which is the total concentration of estrogenic active compounds in an environmental sample normalized to that of natural estrogen E2, determined by yeast estrogen screen bioassay) in one pilot-scale MBR and two full-scale STPs receiving the same domestic wastewater and showed that MBR removes estrogenic activity more effectively than STPs. The MBR showed a 73% removal of estrogenic activity from domestic wastewater, which was 10–15% greater than the two STPs examined [22]. A recent study also reported that the MBR has relatively good performance in the removal of estrogenic activity compared to STP [98]. The overall estrogenic activity of MBR effluents was in the range of 0.75 ± 0.44 to 1.24 ± 0.22 ng EEQ/L for pilot-scale plant and 1.60 ± 0.93 to 1.70 ± 0.56 ng EEQ/L for lab-scale system. Those are comparable to those of other effluents, e.g., 2.9 ng EEQ/L from the STP located in Cobis, Italy [17] and 2.5 ng EEQ/L from Penha, Rio de Janeiro [75], especially lower than the values of 10.5 ng EEQ/L from the STP in Frankfurt/Main, Germany [75], and the 6 ng EEQ/L reported by Körner et al. [21]. The observed decrease in overall estrogenicity in MBR may be caused by its process intensification through high biomass concentration.

Estrogenic activity is removed with degradation and/or transformation of the parent estrogen compound. There have been some studies on the fate and removal efficiency of EDCs in MBR systems. Reports on EDCs removal by MBRs are listed in Table 3. The literature data suggest that MBR has good removal capacities for EDCs, although removal efficiencies differ from different reports. In a study by Joss et al. [68] on removal of E1, E2, and EE2 in various municipal STP and MBR, the results show comparable removal of EDCs in the MBR (96% for E1, >98% for E2, and >75% for EE2) and in three conventional activated sludge systems (49–99% for E1, 88–98% for E2, and 71–94% for EE2). For E1 and E2, degradation activity was seen to be higher in the MBR sludge than in the conventional activated sludge by a factor of 2–3. They explained this difference in activity by the longer sludge age (30 days in MBR vs. 12 days in STP) and smaller floc size (10–100 μm in MBR vs. 100–500 μm in STP).

Water Reclamation System and Micropollutants. Table 3 Summary of some studies on fate and removal efficiencies of EDCs in MBR systems

EDCs	MBR	Operating factor conditions	Removal (%)	Effluent concentration (ng/L)	References
E1	Submerged; MF-UF	SRT = 30 days; MLSS = 0.22–0.29 g/L	96	2.4 ± 0.5	[68]
	External; UF	SRT at 10°C = 22, 40, 82 days; HRT = 0.5–4 d; MLSS = 6.3–11.8 g/L	27.6–100	ND ^a –21	[44]
	Submerged; UF	SRT = 40 day; MLSS = 10–15 g/L	NA ^b	ND–1.2	[99]
E2	Submerged; MF-UF	SRT = 30 day; MLSS = 0.22–0.29 g/L	>98	<0.5	[68]
	External; UF	SRT at 10°C = 22, 40, 82 days; HRT = 0.5–4 d; MLSS = 6.3–11.8 g/L	95.2–100	ND–6	[44]
	Submerged; MF	MLSS = 2.7–3.5 g/L; pH = 4.3–8.0	98–100		[100]
	Submerged; UF	SRT = 40 days; MLSS = 10–15 g/L	NA	ND–1.1	[99]
E3	External; UF	SRT at 10°C = 22, 40, 82 days; HRT = 0.5–4 d; MLSS = 6.3–11.8 g/L	99.7–100	ND–1	[44]
EE2	Submerged; MF-UF	SRT = 30 day; MLSS = 0.22–0.29 g/L	>75	<0.5	[68]
	External; UF	SRT at 20°C = 11, 20, 40 days	60–70	1–3	[101]
	External; UF	SRT at 20°C = 11, 20, 40 days	25–80	NA	[102]
	External; UF	SRT at 10°C = 22, 40, 82 days; HRT = 0.5–4 d; MLSS = 6.3–11.8 g/L	33.3–100	ND–4	[44]
	Submerged; MF	MLSS = 2.7–3.5 g/L; pH = 4.3–8.0	60–70	NA	[100]
	Submerged; UF	SRT = 40 day; MLSS = 10–15 g/L		0.9–1.6	[99]
BPA	External, UF	SRT at 20°C = 11, 20, 40 days	97–99	NA	[102]
	External; UF	SRT at 10°C = 22, 40, 82 days; HRT = 0.5–4 d; MLSS = 6.3–11.8 g/L	92.7–99.3	16–158	[44, 79]
	Submerged; MF	MLSS = 2.7–3.5 g/L; pH = 4.3–8.0	50–100	NA	[100]
	Submerged; MF	HRT = 6 h; MLSS = 13.6 g/L	97.5	1.22	[103]
	Submerged; UF	SRT = 40 day; MLSS = 10–15 g/L	NA	2.5–12.6	[99]
	Submerged; MF	SRT = 350 d; HRT = 4–8 h; MLSS = 5–13 g/L	>93.7	<0.08	[104]
NP	External, UF	SRT = 10, 27, 55 days; HRT = 0.5–4 d; MLSS = 6.3–11.8 g/L	85–91	297–482	[79]
	Submerged; MF	HRT = 6 h; MLSS = 13.6 g/L	>98	<0.88 ng/L	[103]
	Submerged; MF	HRT = 10 h; MLSS = 11–20 g/L	96	<2 ng/L	[104]

^aNot detected^bNot available

In their study, MBR sludge is characterized by higher specific surfaces. Clara et al. [44] also showed that high removal rates (80–100%) could be achieved for E1, E2, E3, EE2, and BPA in MBR systems operating at

different SRTs. They noted that although high removal rates could be achieved in conventional STP operating at high SRT, the MBR allows achieving high SRT within a compact reactor, resulting in higher volumetric

reaction rates as compared to the conventional STP. Zuehlke et al. [105] studied the performance of two MBR pilot plants (hollow fiber polyvinylidene fluoride membrane modules with pore size of 0.2 μm and membrane area of 9 m^2) operated in parallel to a full-scale STP fed with the same raw municipal sewage. More than 90% of E1 and E2 and about 80% of EE2 were removed by the conventional STP, respectively. Removal rates could further be improved by MBR; E1 and E2 were removed up to 99% and EE2 up to 95%. Especially for EE2, in most cases, its concentrations in MBR effluents were reported to be below or approximately 1 ng/L. Using a pilot-scale MBR, Spring et al. [99] found that MBR was more efficient at removing E1, EE2, and BPA to low ng/L levels than STP receiving the same wastewater. Concentrations found in the MBR effluents ranged between ND–1.2 ng/L (0.6 ng/L) for E1, 0.9–1.6 ng/L (1.1 ng/L) for EE2, and 2.5–12.6 ng/L (5.6 ng/L) for BPA, while those in STP effluents were 0.9–2.9 ng/L (1.6 ng/L) for E1, 1.0–2.7 ng/L (1.6 ng/L) for EE2, and 3.5–15.8 ng/L (8.7 ng/L) for BPA, respectively.

Some studies focused on the removal of BPA and NP from wastewater during MBR treatment [103–105, 107]. Both substances were reported to be removed to a large extent by MBR treatment. In a study of Schröder [103], the elimination BPA and NP in MBR treatment was 97.8% and >98%, respectively. Concentrations of BPA and NP observed in the MBR effluents were $1.22 \pm 0.10 \mu\text{g/L}$ and below the detection limit of 0.88 ng/L, while those in the conventional STP effluents were $4.63 \pm 1.51 \mu\text{g/L}$ and $0.92 \pm 0.11 \mu\text{g/L}$, respectively. Chen et al. [104] studied the removal efficiencies of BPA at different BPA sludge loadings. The results showed BPA removal efficiency in MBR higher than 94% and that MBR could effectively remove BPA despite the changes in sludge loadings of BPA ranging from 0.046 to 10.2 g/kg·day. Using a lab-scale MBR, González et al. [105] found that MBR treatment could eliminate nonylphenolic compounds with an overall efficiency of 94%, which represented significant improvement in comparison to the STP treatment where only 54% of the total nonylphenolic compounds were removed. Cirja et al. [108] performed a study aimed at giving a better insight into the possible fate of NP during wastewater treatment by using a lab-scale MBR designed and optimized for fate studies carried

out with radiolabeled compounds. The mass balance of NP residues at the end of the study showed that 42% of the applied radioactivity was recovered in the effluent as degradation products of NP, 21% was removed with the daily excess sludge from the MBR, and 34% was recovered as adsorbed in the component parts of the MBR. A high amount of NP was associated to the sludge during the test period, while degradation compounds were mainly found in the effluent.

Recently, the fate of EDCs and their conjugates was investigated in 3 pilot-scale and 2 lab-scale MBR systems [98]. Pilot-scale MBRs, where the membranes were submerged directly into anoxic and aerobic bioreactors, were at SRT of 21 days and HRT of 6 h. Lab-scale MBRs were operated at different SRTs (17 and 33 days) with schematic diagrams similar to the pilot-scale MBR. The concentrations of estrogens and conjugates, as well as their removal efficiencies, in the various MBR systems are shown in Table 4. The high removal rate of E1 (80.2–91.4%) in MBR systems, compared to $61 \pm 38\%$ [17] and around $\sim 80\%$ [78] in STPs reported in literature, means that E1 could be eliminated very fast, even though E1 can be produced in the MBR systems by oxidation of E2 or deconjugation of estrogen conjugates. While compared with the removal rate of E2 of 64% [75] and 85% [41] by STPs, E2 removal efficiency in this study (49.3–63.1%) was relatively low. The low concentration of conjugates in the MBR effluents, which had a maximum concentration of 27 ng/L, suggested that a majority of steroid conjugates was removed in the MBR process by various bacteria stains in the activated sludge. The fecal bacteria *Escherichia coli*, which exists in large quantities in the feces, is able to synthesize large amounts of the β -glucuronidase enzymes [109], and it has been suggested to be responsible for the removal of estrogen glucuronides after treatment. Furthermore, Johnson and Sumpter [73] suggested that the activated sludge containing arylsulfatase enzyme could achieve desulfation. It is however clear that substantial amounts of E1, E1-3S, E1-3G, and E2-G still passed through the treatment system and entered the aquatic environment. Therefore, it is inadequate to be concerned with only free estrogens in the effluent of treatment plants, as conjugates can be converted to free estrogens further downstream in the receiving water.

Water Reclamation System and Micropollutants. Table 4 Concentrations and removal efficiencies of natural estrogens and conjugates in MBR systems

	E1	E2	E1-3G	E1-3S	E2-G (3G+17G)	E2-3S	E2-3,17S	BPA	4-NP
Influent (ng/L)									
Pilot	16.50–44.05 (27.81)	ND ^a –5.74 (ND)	ND–43.4 (24.7)	24.78–39.03 (29.85)	82.50–97.50 (92.05)	10.68–13.14 (12.03)	ND–99.4 (30.21)	0.01–2.05 (0.18)	ND–1.21 (0.16)
Lab	16.63–49.95 (32.11)	ND–6.21 (ND)	ND–30.10 (20.95)	20.49–25.56 (22.02)	41.02–99.50 (50.00)	8.82–17.70 (13.62)	ND	ND–1.20 (0.15)	ND–9.07 (0.41)
Effluent (ng/L)									
Pilot-A	1.81–11.21 (4.24)	ND–2.06 (ND)	ND–11.02 (6.60)	9.03–17.97 (15.30)	12.23–24.34 (18.05)	ND–5.13 (ND)	ND	ND–0.16 (0.05)	0.002–4.34 (0.64)
Pilot-B	1.71–12.12 (3.25)	ND–2.07 (ND)	ND–11.15 (8.05)	9.15–18.39 (16.41)	17.85–35.25 (27.00)	ND–4.68 (ND)	ND	ND–0.60 (0.07)	0.01–5.93 (0.42)
Pilot-C	1.82–7.21 (4.51)	ND–2.83 (ND)	ND–20.21 (16.30)	10.53–26.25 (23.13)	14.32–28.47 (24.05)	ND–5.46 (ND)	ND	ND–0.07 (0.04)	0.07–6.70 (0.72)
Lab-17d	2.70–3.78 (2.96)	ND	2.50–4.95 (3.10)	17.61–19.71 (17.79)	15.58–26.25 (19.88)	ND–8.22 (ND)	ND	ND–0.25 (0.03)	ND–1.88 (1.12)
Lab-33d	ND–2.83 (2.76)	ND	2.73–4.60 (3.70)	13.53–24.51 (17.80)	13.50–23.33 (17.25)	ND	ND	ND–0.14 (ND)	ND–2.12 (1.41)
Removal (%)									
Pilot-A	80.3	63.1	84.9	38.3	80.4	85.2	100	73.7	–417.6
Pilot-B	82.3	62.9	77.0	48.3	73.5	91.1	100	68.9	–161.1
Pilot-C	80.2	49.3	62.4	10.1	73.9	80.2	100	77.9	–439.5
Lab-17d	84.0	60.0	81.2	21.5	55.1	84.5	–	82.2	–269.3
Lab-33d	91.4	66.5	83.5	32.2	67.2	100	–	90.1	–313.4

Source: Hu et al. [98]

^aNot detected; Figures in brackets indicate median value

It is clear that the removal rates of E1-3G (62.4–84.9%), E2-G (55.1–80.4%), E1-3S (10.1–48.3%), and E2-3S (80.2–100%) were lower than those reported by D'Ascenzo et al. [41], which were 84%, 100%, 64%, and 100%, respectively, in STPs (Table 4). The low removal rates of estrogen conjugates may be caused by two reasons. First, the relatively short HRT of the MBR systems (6 h) compared with STPs did not allow enzymes sufficient time to degrade conjugates in the MBR process, as both glucuronides and sulfates are difficult to convert. Reported half-lives are 6 h for estrogen glucuronides and 2.5 days for estrogen sulfates [41]. Second, all MBRs were operated with pre-denitrification zones, which may have decreased the biodegradation rate of the compounds of concern. Moreover, the extremely low removal rate of E1-3S could be caused by its slow biotransformation and conversion from E2-3S or other parent conjugates.

As shown in Table 4, for alkylphenol compounds, BPA was removed well with a removal efficiency of 68.9–90.1%, but 4-NP concentration was amplified (removal efficiency of –439.5% to –161.1%) in almost all MBR systems, as compared with that of influents. This indicates that 4-NP could be produced from parent compounds in the MBR systems.

In general, an enhanced degree of EDCs removal has been discoverable for MBRs. However, in the comparative study on removal of EDCs by MBR and conventional STP, no significant differences between the removal efficiencies in the conventional STP and the MBR have also been observed. In a study of Lesjean et al. [110], for instance, high removal rates (>80%) could be achieved for less polar EDCs (E1, E2 and EE2), which represents only a slight improvement since the STP treatment commonly yields elimination of 75–95%. The comparative study of Clara et al. [101] also showed similar elimination of EE2 in MBR and STP systems, with removal rates of 60–70%. The EE2 degradation is more likely linked to the activity of nitrifying bacteria present in biological treatment process [111–113]. In a study of Urase et al. [100], the removal of E2 by MBR was always recorded to be high (~100%) regardless of pH, while the removal of BPA was low (~50%) in the acidic pH condition (4.3–5.0). A considerable part of E2 was converted to E1 and a part of converted E1 remained in the MBR effluent. Weber et al. [114] report that E2 turnover rates to E1

did not differ greatly between conventional STP and MBR sludge. For BPA and NP, high elimination rates (>90%) could be achieved in both STP and MBR [79]. Especially in STP, degradation of alkylphenolic compounds appears to be a function of the operated SRT. Clara et al. [79] observed no removal of BPA and NP in the high-loaded STP operating at SRT between 1 and 2 days. The removal rate can vary depending on the operating characteristics of STP (i.e., plant size, SRT, HRT, temperature, influent concentration, and adaptation time), although high removal rate can be achieved in a well-operated STP. On the other hand, despite achieving a better elimination, operating problems such as bad sludge characteristics and foam formation in MBR systems can affect the variation in effluent concentrations of EDCs [115]. However, MBR processes could cope much better with the fluctuations in the influent concentrations of EDCs, despite the low sludge age [106]. Furthermore, due to the higher diversity of microorganisms in MBR, the acclimating procedure of sludge to EDCs such as BPA in MBR would require less time than in conventional STP [104].

MBR is generally shown to be more effective in removing EDCs than conventional STP, although contradictory results are documented in literature. Therefore, the MBR gives room for expectation that with this alternative wastewater reclamation, the environmental risk associated with EDCs can be reduced. This is probably because, first, MBR is operated under longer SRT that may facilitate additional biological transformation of EDCs. Second, because of the higher biomass concentrations, MBR is operated under the low F/M ratio, which could result in more complete mineralization of poorly degradable organic compounds. Third, the membrane (MF or UF) serves as a complete barrier to the estrogen-bound solids or colloids. Finally, an additional removal of hydrophobic compounds by membranes can be expected, because they are sorbed onto the particles that form a deposition layer on the membrane surface. These characteristics are generally not provided in STPs.

Mechanisms of EDCs Removal

The mechanisms of EDCs removal is currently one important focus of research, because several mechanisms may exist through which MBRs achieve a better

elimination of EDCs than STP. Considering low values of vapor pressure of most of the EDCs, stripped fraction removed by volatilization can be neglected [85]. Since MF and UF membranes applied in MBR have a pore size between 100 and 10,000 times larger than the physical size of EDCs (molecular weight lower than 300 g/mol), direct physical retention by the membranes is of minor for EDCs. In MBR, EDCs can sorb to solids and subsequently be separated from the treated water by MF or UF. In addition, EDCs can be biodegraded after an association with the biomass in the MBR. Therefore, the removal of EDCs in MBR system is commonly achieved by sorption on solids, biodegradation, or a combination of both.

Sorption Estrogens and their conjugates can be sorbed to biomass before appreciable degradation occurs. Andersen et al. [116] reported that the sum of sorbed E1 and E2 to sludge ranged from 7.8 to 13 µg/kg. Ternes et al. [117] also reported that up to 37 µg/kg of E1 and 49 µg/kg of E2 sorbed onto sludge. In a study for the evaluation of EDCs on MBR sludge [98], the levels of sorbed E1, E1-3S, E2-3S, BPA, and 4-NP were up to 10.42, 18.32, 2.52, 34,350, and 68,800 µg/kg,

respectively (Table 5). However, no estrogen glucuronides could be detected on MBR sludge. Although natural estrogen sulfates are relatively hydrophilic, the sorbed amounts are comparable to that of E1. It can be seen that generally the amount of 4-NP sorbed onto the sludge solids is more than that of BPA sorbed onto the sludge solids (Table 5). This may be due to the fact that 4-NP has a higher log K_{OW} and lower aqueous solubility, so 4-NP is more likely to be adsorbed onto sludge solids.

Sorption of organic micropollutants to the sludge in treatment plants depends on two main mechanisms [118]: absorption and adsorption. Absorption is the hydrophobic interactions of the aliphatic and aromatic groups of a compound with the lipophilic cell membrane of the microorganisms and with the lipid fraction of the sludge. Adsorption is the electrostatic interactions of positively charged groups of the chemicals with the negatively charged surfaces of the biomass. Absorption is related to the substance's lipophilicity, characterized by the K_{OW} . Adsorption is related to the tendency of a substance to be ionized or dissociated in aqueous phase, which is characterized by the dissociation constant (pK_a).

Water Reclamation System and Micropollutants. Table 5 Amounts of natural estrogens and conjugates sorbed to the sludge in MBR systems

	E1	E1-3S	E2-3S	BPA	4-NP
Anoxic (µg/kg)					
Pilot-A	ND ^a -9.70 (6.80)	ND-11.91 (3.44)	ND-0.31 (0.22)	0.47-14,460 (2,070)	7,245-13,600 (10,220)
Pilot-B	ND-6.40 (5.64)	ND-14.68 (4.36)	ND-1.92 (0.97)	0.55-34,350 (2,620)	ND-4,610 (3,010)
Pilot-C	ND-8.11 (5.62)	ND-11.21 (3.48)	ND-0.23 (0.16)	0.01-8,480 (4,250)	ND-4,600 (3,627)
Lab-17d	ND-7.66 (4.98)	ND-2.48 (1.11)	ND-0.28 (0.12)	1.04-8,460 (1,570)	ND-6,660 (2,800)
Lab-33d	ND-8.40 (7.38)	ND-5.78 (2.35)	ND-1.22 (0.42)	1.47-10,560 (2,080)	ND-11,400 (4,000)
Aerobic (µg/kg)					
Pilot-A	ND-2.04 (1.40)	ND-6.21 (4.52)	ND-0.75 (0.38)	2.24-1,820 (353.38)	3020-15,630 (13,380)
Pilot-B	ND-3.73 (1.56)	ND-8.36 (5.16)	ND-2.52 (1.13)	0.89-2,000 (221.13)	278.15-68,800 (16,960)
Pilot-C	ND-4.42 (2.24)	ND-8.21 (4.84)	ND-0.34 (0.18)	0.03-1,360 (273.22)	290.16-56,700 (10,490)
Lab-17d	ND-2.42 (1.22)	ND-5.32 (2.77)	ND-0.22 (0.12)	ND-1,190 (157.23)	ND-9,300 (5,160)
Lab-33d	ND-10.42 (7.98)	ND-10.96 (6.21)	ND-1.96 (0.82)	ND-1,250 (189.19)	134.02-18,130 (10,060)

Source: Hu et al. [98]

^aNot detected; Figures in brackets indicate median value

The K_{OW} value reflects the equilibrium of partitioning the organic solute between the organic phase, i.e., octanol and water phases. High K_{OW} is characteristic for hydrophobic compounds, poor hydrosolubility, and high tendency to sorb on organic material of the sludge matrix [85]. Based on the log K_{OW} , sorption to sludge is predicted to play an important role in the removal of hydrophobic compounds from the liquid phase. The following may be used as a general guide [85]: (1) $\log K_{OW} < 2.5$ results in low sorption potential; (2) $2.5 < \log K_{OW} < 4.0$ yields medium sorption potential; (3) $\log K_{OW} > 4.0$ leads to high sorption potential. Generally, the more hydrophobic a chemical substance, the greater the amount accumulated in the solid phase, whereas the more hydrophilic the chemical, the greater its concentration in the aqueous phase. The estrogenic compounds are generally characterized by relatively medium hydrophobicity with typical K_{OW} values between 3 and 5 (Table 1). The synthetic estrogens, due to their higher K_{OW} can be more easily removed from the water phase than the natural estrogens that are characterized by lower hydrophobicities. Apart from this, the presence of other hydrophobic chemicals present in different environments may compete with the estrogens for available binding sites in the sorbent [119].

A common approach to determine the fraction of EDCs sorbed onto sludge is the use of the solid-water distribution coefficient (K_D , in L/kg), defined as the ratio between the concentrations in the solid and liquid phases at equilibrium conditions (Eq. 1).

$$K_D = \frac{C_{\text{sorbed}}}{SS \cdot C_{\text{soluble}}} \quad (1)$$

where C_{sorbed} is the sorbed estrogen concentration onto sludge ($\mu\text{g/L}$), C_{soluble} is the soluble concentration in equilibrium with the sorbed concentration, and SS is the sludge concentration (kg/L). Other characteristic parameters used to describe the adsorption capacity of solids are K_{OM} and K_{OC} values. The K_{OM} (L/kg) relates the K_D value to the organic matter of the sorbent (Eq. 2), whereas the K_{OC} (L/kg) refers to the organic carbon content of the sorbent (Eq. 3).

$$K_{OM} = K_D \cdot \frac{SS}{VSS} \quad (2)$$

where VSS is the volatile sludge concentration (kg/L).

$$K_D = K_{OC} \cdot f_{OC} \quad (3)$$

where f_{OC} is the fraction of organic carbon in the solids.

With mainly hydrophobic interactions, the K_D can be estimated from the K_{OW} (or even better from K_{OC}), whereas by electrostatic interactions, K_D is found empirically. The $\log K_D$ (or $\log K_{OC}$) values of several EDCs appear to be independent of $\log K_{OW}$, indicating that hydrophobic interaction is not the predominant sorption mechanism for EDCs [120]. Besides hydrophobic processes that are taking place, a number of other reactions like complex formations with metal ions, ion exchange, and hydrogen bindings decide about the partition of the organic compound between the solid and the liquid phase [121]. Since the organic fraction of the sludge (i.e., VSS) is relevant for adsorption processes [122], MBR sludge can be expected to have higher sorption potential as the organic matter content is higher compared to the sludge of conventional activated sludge process (CASP). On the other hand, both electrostatic and hydrophobic interactions can govern estrogens and their conjugate adsorption, depending on the properties of the adsorbate such as the surface charge and the existence of hydrophobic domains [123, 124].

Clara et al. [122] investigated sorption behavior of E2, EE2, and BPA to sludge. High adsorption potential to sewage sludge could be observed for these substances, with no saturation levels detected. Substances with $\log K_D$ values lower than 2 are negligible for adsorption. For substances with $\log K_D$ values higher than 4, sorption to the sludge is a major removal process. In the work by Clara et al. [122], all investigated substances show K_D values between these boundaries and proof to be relevant for this removal pathway. Authors also found that the sorption isotherm is linear over a concentration range of several orders of magnitude. Andersen et al. [125] determined the sorption of E1, E2, and EE2 to activated sludge. The experimentally obtained $\log K_D$ values were 2.60 (2.44–2.72, 2.68 (2.45–2.83), and 2.77 (2.65–2.86) L/kg for E1, E2, and EE2, respectively. Using the experimental f_{OC} ($27.7 \pm 0.1\%$) and K_D for the sludge, $\log K_{OC}$ were also determined to be 3.16, 3.24, and 3.32 L/kg for E1, E2, and EE2, respectively. The results for $\log K_{OC}$ by Andersen et al. [125] are consistent with the data determined by Clara et al. [122]. In the work by Ternes et al. [126],

$\log K_D$ for EE2 was determined to be 2.54 (2.49–2.58) L/kg. This corresponds to $\log K_{OC}$ of 2.93 (2.86–3.00) L/kg for EE2. In another study, the specific sorption coefficients K_D , K_{OM} , and K_{OC} of activated sludge for E1, E2, and their conjugates were determined [127]. The mean values of $\log K_D$ were 2.78, 2.61, 1.67, and 1.94 L/kg; $\log K_{OM}$ were 2.96, 2.79, 1.77, and 2.04 L/kg, and those of K_{OC} were 3.31, 3.12, 2.21, and 2.46 L/kg for E1, E2, E1-3S, and E2-3S, respectively. The sorption potential of sludge for a certain estrogen in MBR appears to be similar to that of conventional activated sludge. In a study of Chen et al. [104], the range of $\log K_D$ for BPA in MBR sludge was from 2.00 to 2.75 L/kg, which was similar to the reported value between 1.89 and 2.41 [122]. Authors also concluded that although BPA is hydrophobic, the sorption process contributed very little to the BPA removal.

The solution pH affects the solution chemistry of the estrogens and estrogen conjugates, the activity of the functional groups (carboxylate, phosphate, and amino groups) in the biomass as well as the competition of compounds for the binding site. The effect of pH on adsorption of natural estrogens and their conjugates by the activated sludge is shown in Table 6 [127]. The adsorption isotherm of steroid estrogens and their conjugates by the activated sludge can be described by the Freundlich isotherm. The pH value appears to be an important parameter, which affects the adsorption capacity. In the work by Chen and Hu [127], the sorption capacities of E1 and E2 to activated sludge were the highest at neutral pH, but decreased while pH decreased or increased. For both E1-3S and

E2-3S, sorption capacities were highest at pH 2, while lowest at pH 11.5. The lower capacities of E1 and E2 at pH 2 and 11.5 could be due to the competition of sludge adsorption sites by cations or electrostatic repulsion from particles of similar charges. The lowest adsorption capacity of estrogen conjugates at pH 11.5 could be caused by electrostatic repulsion, and the highest capacity at pH 2 might be due to the increased sulfate adsorbability. In another study, pH conditions irrelevant for the sorption of E1, E2, EE2, and BPA were found [128]. The change in pH from 7.0 to 4.4 has no effect on their K_D [128]. On the other hand, at high pH values, a release of the adsorbed fraction to the liquid phase was observed for E2, EE2, and BPA in a study of Clara et al. [122]. Desorption occurred at pH values in the range of the pKa values (10.4 for E2, 10.7 for EE2, and 10.3 for BPA) of the investigated compounds.

Experimental data on the sorption behavior of EDCs to MBR sludge are very scarce, probably due to the difficulties of solid samples analysis. To overcome this problem, the use of K_D values appears to be a useful tool to predict distribution between both phases. However, an accurate determination of this coefficient is required to avoid misunderstandings in the establishment of the mechanism responsible for the elimination of a particular compound.

Biodegradation Biodegradation is an important estrogen removal mechanism in both STP and MBR systems. Though the degradation pathways are still incomplete, estrogenic activity is removed with

Water Reclamation System and Micropollutants. Table 6 Effect of pH on isothermal constants ($K_F[\text{mg}^{1-1/n} (\text{m}^3)^{1/n} \text{g TSS}^{-1}]$ and $1/n$) of Freundlich model for steroid estrogens and their conjugates

pH	E1		E2		E1-3S		E2-3S	
	K_F	$1/n$	K_F	$1/n$	K_F	$1/n$	K_F	$1/n$
2	0.23	0.94	0.11	1.04	0.189	0.863	0.347	0.881
5	0.26	1.08	0.22	0.96	0.032	0.857	0.064	0.830
7	0.36	1.05	0.60	0.99	0.022	0.739	0.052	0.736
9	0.29	0.92	0.26	0.95	0.032	0.810	0.056	0.807
11.5	0.05	1.00	0.04	1.08	0.006	0.742	0.012	0.887

Source: Chen and Hu [127]

biotransformation of the parent estrogen compound. This deserves further attention when considering the fact that sorption only involves a phase shift of the pollutants and no degradation or even transformation. For nonpolar trace organics, in general, sorption to the biomass and subsequent retention of the solids by the membranes will be a major removal mechanism. For polar compounds, sorption will be limited and elimination can only be achieved through biodegradation.

The biodegradation of micropollutants such as EDCs in the sludge phase is described by pseudo-first-order kinetic model as given below [68, 129]:

$$\frac{dC_T}{dt} = -k_b X_{SS} C \quad (3)$$

where dC_T/dt is the reaction rate (ng/L·day), C_T is the total estrogen concentration (ng/L), k_b (L/g·day) is the pseudo-first-order rate coefficient for estrogen biodegradation, X_{SS} is the concentration of suspended solids (g/L), and C is the soluble estrogen concentration (ng/L). The term “first order” refers to the direct proportionality of the transformation rate to the soluble substance concentration and the term “pseudo” refers to its proportionality to the sludge concentration [129].

Joss et al. [129] have determined biodegradation kinetic constants (k_b) for a large number of compounds with the sludge from CASP and MBR. Based on the k_b values, three groups of compounds were defined according to their susceptibility to biological degradation: (1) hardly biodegradable with $k_b < 0.1$ L/g·day; (2) moderate biodegradable with $0.1 < k_b < 10$ L/g·day; and (3) highly biodegradable compounds with $k_b > 10$ L/g·day. The reported k_b values of E1, E2, and EE2 are in the range of 200–300, 300–800, and 7–9 L/g·day, respectively [129, 130]. From the classification scheme, E1 and E2 are expected to be highly biodegraded, whereas partial biodegradation is predicted for EE2.

In general, trace organic removal is better in MBR than in STP. Yet, the most effective treatment option depends strongly on the target compounds present in wastewater. In treatment process with a combination of different mechanisms (sorption and biodegradation) for removal of compounds, the following considerations can be taken into account

according to the k_b and K_D (or K_{OC}) values of a particular compound [130]:

- Compounds with high k_b and low K_D (or K_{OC}) values are very well transformed independently of SRT and HRT.
- Compounds with low k_b and high K_D (or K_{OC}) values are retained in the aerobic reactor by sorption and significantly transformed when the SRT is long enough to accomplish biological degradation.
- Compounds with high k_b and medium K_D (or K_{OC}) values are moderately transformed independently of HRT and slightly dependant on SRT.
- Compounds with low k_b and K_D (or K_{OC}) values are not removed nor biotransformed regardless of HRT and SRT.

For instance, E1 and E2 belong to compounds with high k_b and medium K_{OC} , while EE2 has low k_b and high K_{OC} values [68, 131]. Clara et al. [44] showed a dependency of the removal of E1 and E2 on SRT in STP and MBR, high removal rates at the prolonged SRT. EE2 could be removed at significant rate under aerobic conditions at sufficient SRT [68]. Therefore, this information would be useful for selection of the treatment option for EDCs removal from wastewaters.

The sludge characteristics are an important factor for biodegradation. There are important sludge characteristics that are to be considered for biological transformation of estrogens [129]: (1) diversity of the activity of the biomass due to either differences in microbial population or the expressed enzymatic activity, (2) floc size of the sludge for compounds being well degraded, and (3) the fraction of active biomass within the total suspended solids. Significant differences in sludge characteristics have been observed between CASP and MBR sludges [44, 47, 68, 93, 94, 129].

In the case of MBR, the bioreactor and membrane units cannot be regarded as individual unit operations because they interact with respect to both biological reactions and membrane filtration. High SRT and full retention of solid particles on membranes enable biological adaptation to less biodegradable compounds and can promote a wider diversity of the microbiological communities, resulting in changes in removal rates and efficiencies. Possible changes in the activated sludge composition can be used for the establishment of “specialized” microorganisms [65]. The membrane

not only retains all biomass but also prevents the escape of exocellular enzymes and soluble oxidants creating a more active biological mixture capable of degrading a wider range of organic compounds [132]. In a study by Cicek et al. [132], enzymatic activity tests showed that (1) overall activity was consistently higher in the MBR, (2) both systems produced enzymes specific for the degradation of feed chemicals, (3) no enzymes were detected in the MBR effluent compared to significant amounts in the CASP, and (4) the MBR contained more enzymes in the soluble phase than the CASP. Data from literature show that higher sludge age, often reached in MBRs, may significantly improve the removal of specific compounds [133, 134]. For compounds like trimethoprim, and macrolide antibiotics azithromycin, erithromycin, and clarithromycin, a clear increase in transformation was found at sludge ages of 60–80 days [134].

A high SRT in a biological process positively affects the overall activity of slow growing microorganisms (e.g., nitrification). Baek and Pagilla [135] found that the high SRT (103 ± 9 days) of MBR system retained more active nitrifying bacteria than CASP (SRT of 10–15 days). Furthermore, MBR enables nitrifying bacteria to develop and persist in the system even under short SRTs (2–10 days) [96]. Complete nitrification was demonstrated in MBRs at sludge ages of 5–72 days [136]. The treatment process configuration may affect the removal of estrogens as has been shown by significant higher removal rates with nitrifying activated sludge [76]. Estrogens removal rates of greater than 90% were achieved in MBRs with nitrification and denitrification [68]. Vader et al. [111] demonstrated the degradation of EE2 by a nitrifying sludge with a high ammonium (NH_4^+) oxidizing activity. Shi et al. [137] and Yi and Harper [138] used nitrifying mixed cultures to degrade EE2 and suggested that the key enzyme for nitrification, i.e., ammonium monooxygenase, could mediate an EE2/ NH_4^+ cometabolism. These authors brought forward the concept that nitrifiers initially degrade EE2 into intermediates that subsequently serve as a substrate for heterotrophic organisms. On the other hand, some authors suggested that the primary mechanism for EE2 degradation is more likely linked to the activity of heterotrophic bacteria [139]. They suggested that the EE2 removal by axenic cultures of ammonia-oxidizing bacteria is only due to the

abiotic nitrification reaction with EE2, governed by the high NO_2^- -N levels after oxidation of the high NH_4^+ -N concentrations in batch tests.

The biomass developed in MBR systems can be characterized by smaller floc size with respect to the removal efficiency of micropollutants. In CASP, flocs may reach several 100 μm in size [140]. This means that the substrate can reach the active sites inside these flocs only by diffusion, which causes an additional resistance and limits the overall reaction rate. Hydrodynamic stress in MBRs reduces floc size to approximately 30–60 μm in immersed systems [92, 141] and 3.5 μm in side-stream MBRs [94], and thereby, indirectly increases the apparent reaction rate. Smaller flocs can explain enhanced mass transfer for both oxygen and carbon, thus enabling a higher removal rate and more adaptability to changes in the influent quality and quantity [142]. In another study, it was demonstrated that the flocs were more active and displayed greater species diversity [132]. Furthermore, some enzymatic activities increase proportionally to the higher specific surface area of MLSS, which is directly related to the floc structure [85].

SRT has been reported as determinative for biodegradation of micropollutants. Clara et al. [44] suggested that a critical value for the sludge age can be determined. With regard to natural estrogens, a strong correlation between achievable effluent concentrations and the $\text{SRT}_{10^\circ\text{C}}$ (temperature corrected SRT) was observed [44]. For these substances, a critical value was identifiable for the $\text{SRT}_{10^\circ\text{C}}$ amounting to approximately 10 days, but for EE2, contradictory results were obtained and beside the SRT, other influences seem to be of importance [44]. These authors concluded that for STPs operated at $\text{SRT}_{10^\circ\text{C}}$ higher than 10 days, low effluent concentrations can be achieved for many biodegradable micropollutants, which corresponds to an SRT required for efficient nitrogen removal. On the other hand, at long SRT, the accumulation of dead or inactive microorganisms occurs in the MBR, and thus affects the sludge activity [93, 143]. Joss et al. [129] noted that the slightly lower transformation rate constants found in MBR operated at 30–40 days of SRT as compared to CASP sludge with 10–12 days of SRT operation, may be, caused by the higher inert matter accumulation in the MBR sludge. Baek and Pagilla [135] suggested that the biomass in the MBR system

with longer SRT of 103 ± 9 days was not in a physiological state characteristic for growth due to low substrate per unit biomass. Increasing the SRT beyond 30 days does not seem to improve the removal for micropollutants [130]. An explanation could be in the fact that biodegradation of micropollutants is most likely performed in a cometabolic manner as the low concentrations do not likely sustain growth for specific microorganisms, because in this case, the SRT necessary for an efficient biodegradation of the primary substrate is the relevant parameter [144].

Several researchers have studied the influence of initial concentration of EDCs and MLSS concentration on EDCs biodegradation performance. Ternes et al. [77] reported the behavior and occurrence of E2 and E2 glucuronides in aerobic batch experiments with activated sludge, and they found that E2 and E2 glucuronides in the ng/mL range were cleaved and eliminated even faster than in the experiments spiked with higher concentrations (mg/mL range). More recently, Li et al. [145] studied the degradation kinetics of E2 under different operating parameters in aerobic batch experiments by activated sludge, and they found comparably that the degradation rate with higher initial concentration was faster by making use of first-order model. In another study, it was demonstrated that the biodegradation of E2 and its conjugates depends on the initial concentration and MLSS concentration of activated sludge [127]. There was a clear trend that the degradation rates of E2, E2-3G, and E2-3S generally increased with the increase of initial concentration range from 10 to 200 $\mu\text{g/L}$ at MLSS of 4 g/L. It suggests that the higher initial substrate concentration has more stimulatory effect on the microorganisms in activated sludge. This could be explained by the reason that the degradation intermediates of E2 or its conjugates did not compete for the degradation by relatively high biomass concentration (4 g/L) with studied compound itself, and the inhibitory effect was unnoticeable. The same tendency was also reported by Li et al. [145], in which they used sludge (MLVSS of 1.75 g/L) to degrade E2 with initial concentration of 10, 30, and 50 $\mu\text{g/L}$. Chen and Hu [127] also found that MLSS concentration has a great effect on the degradation rates of E2 and its conjugates. Results showed that the biodegradation rates of E2 in the initial concentration of 50 $\mu\text{g/L}$ increased with the increase of MLSS concentration of

activated sludge from 4 to 12.5 g/L. This could be explained by more microorganisms and bacterial population in activated sludge at higher MLSS concentration that would carry out the degradation of E2 and its conjugates more effectively.

Future Directions

The reclamation and reuse of treated water imposes as the most adequate solution for the future sustainable water management. One of the key issues in water recycling is the emerging problem of micropollutants such as EDCs. Conventional wastewater treatment plants are not effective in the complete elimination of EDCs and their discharges have a great potential to affect the aquatic environment. So far, several studies have proved that MBR is able to achieve an enhanced degree of EDCs removal from wastewater compared to conventional wastewater treatment. However, most of research data are still incomplete or inconclusive on: (1) the predominant removal mechanism responsible for the respective EDC removal under real MBR process conditions, and (2) the understanding of biotransformation pathways of EDCs and their metabolites with estrogenic activity during MBR treatment. Within these areas, several research topics can be defined. An emerging task in MBR process treating EDCs present in wastewater is the complete mineralization associated to negligible effluent concentrations. It can be achieved through the optimization of determinative parameters (e.g., SRT, HRT, organic loading, sludge concentration, redox conditions, and reactor configuration, etc.) for EDCs removal.

Research into identification of microbial species and strains that can degrade estrogens with the possible application to wastewater treatment has led researchers to explore other areas. For instance, using strains of *Rhodococcus* isolated in activated sludge from wastewater treatment plants, Yoshimoto et al. [146] found rapid degradation of high concentration (100 mg/L) of natural and synthetic steroid estrogens (E1, E2, E3, and EE2). Therefore, search for specific microorganisms with high estrogen-degrading activity and their use in wastewater treatment deserves further research. Introducing estrogen-degrading strains into MBR is an interesting option in a way that would provide a flexible and efficient degradation.

Bibliography

- Colborn T, Dumanoski D, Myers JP (1996) Our stolen future: Are we threatening our fertility, intelligence, and survival? – A scientific detective story. Dutton, New York
- Reddy S, Iden CR, Brownawell BJ (2005) Analysis of steroid conjugates in sewage influent and effluent by liquid chromatography–tandem mass spectrometry. *Anal Chem* 77:7032–7038
- Harris CA, Henttu P, Parker MG, Sumpter JP (1997) The estrogenic activity of phthalate esters *in vitro*. *Environ Health Perspect* 105:802–811
- Desbrow C, Routledge EJ, Brighty GC, Sumpter JP, Waldock M (1998) Identification of estrogenic chemicals in STW effluent 1. Chemical fractionation and *in vitro* biological screening. *Environ Sci Technol* 32:1549–1558
- Jobling S, Nolan M, Tyler CR, Brighty G, Sumpter JP (1998) Widespread sexual disruption in wild fish. *Environ Sci Technol* 32:2498–2506
- Preziosi P (1998) Natural and anthropogenic environmental oestrogens: the scientific basis for risk assessment endocrine disrupters as environmental signalers: An introduction. *Pure Appl Chem* 70:1617–1631
- Holm L, Berg C, Brunstrom B, Ridderstrale Y, Brandt I (2001) Disrupted carbonic anhydrase distribution in the avian shell gland following *in ovo* exposure to estrogen. *Arch Toxicol* 75:362–368
- Chilvers C, Pike MC, Forman D, Fogelman K, Wadsworth ME (1984) Apparent doubling of frequency of undescended testicles in England and Wales 1962–81. *Lancet* 2:330–332
- Kimmel CA (1993) Approaches to evaluating reproductive hazards and risks. *Environ Health Perspect* 101:137–143
- Rajpert-De Meyts E, Skakkeboek NE (1993) The possible role of sex hormones in the development of testicular cancer. *Eur Urol* 23:54–61
- UK Environment Agency (1997) The identification and assessment of oestrogenic substances in sewage treatment works effluents. Bristol, UK
- Purdom CE, Hardiman PA, Bye VJ, Eno NC, Tyler CR, Sumpter JP (1994) Oestrogenic effects of effluents from sewage treatment works. *Chem Ecol* 8:275–285
- Routledge EJ, Sheahan D, Desbrow C, Brighty GC, Waldock M, Sumpter JP (1998) Identification of estrogenic chemicals in STW effluent. 2. *In vivo* responses in trout and roach. *Environ Sci Technol* 32:1559–1565
- Jobling S, Sheahan D, Osborne JA, Matthiessen P, Sumpter JP (1996) Inhibition of testicular growth in rainbow trout (*Oncorhynchus mykiss*) exposed to estrogenic alkylphenolic chemicals. *Environ Toxicol Chem* 15:194–202
- Länge R, Hutchinson TH, Croudace C, Siegmund F, Schweinfurth H, Hampe P, Panter GH, Sumpter JP (2001) Effects of the synthetic estrogen 17 α -ethinylestradiol on the life-cycle of the fathead minnow (*Pimephales promelas*). *Environ Toxicol Chem* 20:1216–1227
- Jobling S, Sumpter JP (1993) Detergent components in sewage effluent are weakly oestrogenic to fish: An *in vitro* study using rainbow trout (*Oncorhynchus mykiss*) hepatocytes. *Aquat Toxicol* 27:361–372
- Baronti C, Curini R, D'Ascenzo G, Di Corcia A, Gentili A, Samperi R (2000) Monitoring natural and synthetic estrogens at activated sludge sewage treatment plants and in a receiving river water. *Environ Sci Technol* 34:5059–5066
- Huang CH, Sedlak DL (2001) Analysis of estrogenic hormones in municipal wastewater effluent and surface water using enzyme-linked immunosorbent assay and gas chromatography/tandem mass spectrometry. *Environ Toxicol Chem* 20:133–139
- Snyder S, Villeneuve DL, Snyder E, Giesy JP (2001) Identification and quantification of estrogen receptor agonists in wastewater effluents. *Environ Sci Technol* 35:3620–3625
- Spengler P, Körner W, Metzger JW (2001) Substances with estrogenic activity in effluents of sewage treatment plants in southwestern Germany. 1. Chemical analysis. *Environ Toxicol Chem* 20:2133–2141
- Körner W, Bolz U, Submuth W, Hiller G, Schuller W, Hanf V, Hagenmaier H (2000) Input/output balance of estrogenic active compounds in a major municipal sewage plant in Germany. *Chemosphere* 40:1131–1142
- Holbrook RD, Novak JT, Grizzard TJ, Love NG (2002) Estrogen receptor agonist fate during wastewater and biosolids treatment process: a mass balance analysis. *Environ Sci Technol* 36:4533–4539
- Smith CW, Di Gregorio D, Talcott RM (1969) The use of ultrafiltration membrane for activated sludge separation. In: Proceedings of the 24th annual Purdue industrial waste conference. Purdue University, West Lafayette, pp 1300–1310
- Stephenson T, Judd S, Jefferson B, Brindle K (2000) Membrane bioreactors for wastewater treatment. IWA, London, UK
- Visvanathan C, Ben Aim R, Parameshwaran K (2000) Membrane separation bioreactors for wastewater treatment. *Crit Rev Environ Sci Technol* 30(1):1–48
- Teske SS, Arnold RG (2008) Removal of natural and xenoestrogens during conventional wastewater treatment. *Rev Environ Sci Biotechnol* 7:107–124
- Matsui S, Takigami H, Matsuda T, Taniguchi N, Adachi J, Kawami H, Shimizu Y (2000) Estrogen and estrogens mimics contamination in water and the role of sewage treatment. *Water Sci Technol* 42(12):173–179
- Christiansen BL, Winther-Nielsen M, Helweg C (2002) Feminization of fish: the effect of estrogenic compounds and their fate in sewage treatment plants and nature. Environmental project no. 729, Danish Environmental Protection Agency, Denmark, 184 pp
- Jürgens MD, Holthaus KIE, Johnson AC, Smith JLL, Hetheridge M, Williams RJ (2002) The potential for estradiol and ethinylestradiol degradation in English rivers. *Environ Toxicol Chem* 21:480–488
- López de Alda MJ, Barceló D (2001) Review of analytical methods for the determination of estrogens and progestogens in wastewaters. *Fresenius J Anal Chem* 371:437–447

31. Li MH, Wang ZR (2005) Effect of nonylphenol on plasma vitellogenin of male adult guppies (*Poecilia reticulata*). *Environ Toxicol* 20:53–59
32. Nozaka T, Abe T, Matsuura T, Sakamoto T, Nakano N, Maeda M, Kobayashi K (2004) Development of vitellogenin assay for endocrine disrupters using medaka (*Oryzias latipes*). *Environ Sci* 11:99–121
33. Paulozzi LJ (1999) International trends in rates of hypospadias and cryptorchidism. *Environ Health Perspect* 107:297–302
34. Krishnan AV, Stathis P, Permuth SF, Tokes L, Feldman D (1993) Bisphenol A: An estrogenic substance is released from polycarbonate flasks during autoclaving. *Endocrinology* 132:2279–2286
35. Gaido KW, Leonard LS, Lovell S, Gould JC, Babai D, Portier CJ, McDonnell DP (1997) Evaluation of chemicals with endocrine modulating activity in a yeast-based steroid hormone receptor gene transcription assay. *Toxicol Appl Pharmacol* 143:205–212
36. Kuiper G, Carlson B, Grandien K, Enmark E, Haggblad J, Nilsson S, Gustafsson JA (1997) Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptor α and β . *Endocrinology* 132:2279–2286
37. Routledge EJ, Sumpter JP (1996) Estrogenic activity of surfactants and some of their degradation products assessed using a recombinant yeast screen. *Environ Toxicol Chem* 15:241–248
38. Belfroid AC, Van der Horst A, Vethaak AD, Schäfer AJ, Rijs GBJ, Wegener J (1999) Analysis and occurrence of estrogenic hormones and their glucuronides in surface water and waste water in The Netherlands. *Sci Total Environ* 225:101–108
39. Swartz CH, Reddy S, Benotti SJ, Yin H, Barber LB, Brownawell BJ, Rudel RA (2006) Steroid estrogens, nonylphenol ethoxylate metabolites and other wastewater contaminants in ground water affected by a residential septic system on Cape Cod, MA. *Environ Sci Technol* 40:4894–4902
40. Kuch HM, Ballschmiter K (2001) Determination of endocrine-disrupting phenolic compounds and estrogens in surface and drinking water by HRGC-(NCI)-MS in the pictogram per liter range. *Environ Sci Technol* 35:3201–3206
41. D'Ascenzo G, Corcia AD, Gentili A, Mancini R, Nazzari RMM, Samperi R (2003) Fate of natural estrogen conjugates in municipal sewage transport and treatment facilities. *Sci Total Environ* 302:199–209
42. Isobe T, Shiraishi H, Yasuda M, Shinoda A, Suzuki H, Morita M (2003) Determination of estrogens and their conjugates in water using solid-phase extraction followed by liquid chromatography-tandem mass spectrometry. *J Chromatogr A* 984:195–202
43. Fine DD, Breidenbach GP, Price TL, Hutchins SR (2003) Quantitation of estrogens in ground water and swine lagoon samples using solid-phase extraction, pentafluorobenzyl/trimethylsilyl derivatizations and gas chromatography-negative ion chemical ionization tandem mass spectrometry. *J Chromatogr A* 1017:167–185
44. Clara M, Kreuzinger N, Strenn B, Gans O, Kroiss H (2005) The solids retention time—a suitable design parameter to evaluate the capacity of wastewater treatment plants to remove micropollutants. *Water Res* 39:97–106
45. Vethaak AD, Lahr J, Schrap SM, Belfroid AC, Rijs GB, Gerritsen A (2005) An integrated assessment of estrogenic contamination and biological effects in the aquatic environment of the Netherlands. *Chemosphere* 59:511–524
46. Hohenblum P, Gans O, Moche W, Scharf S, Lorbeer G (2004) Monitoring of selected estrogenic hormones and industrial chemicals in groundwaters and surface waters in Austria. *Sci Total Environ* 333:185–193
47. Servos MR, Bennie DT, Burnison BK, Jurkovic A, McInnis R, Neheli T, Schnell A, Seto P, Smyth SA, Ternes TA (2005) Distribution of estrogens, 17 β -estradiol and estrone, in Canadian municipal wastewater treatment plants. *Sci Total Environ* 336:155–170
48. Laganà A, Bacaloni A, Leva ID, Faberi A, Fago G, Marino A (2004) Analytical methodologies for determining the occurrence of endocrine disrupting chemicals in sewage treatment plants and natural waters. *Anal Chim Acta* 501:79–88
49. Kolok AS, Snow DD, Kohno S, Sellin MK, Guillelte JLJ (2007) Occurrence and biological effect of exogenous steroids in the Elkhorn river, Nebraska, USA. *Sci Total Environ* 388:104–115
50. Tan BL (2006) Chemical and biological analyses of selected endocrine disruptors in wastewater treatment plants in south east Queensland, Australia. Doctor dissertation. Griffith University, Queensland
51. Kolodziej EP, Sedlak DL (2007) Rangeland grazing as a source of steroid hormones to surface waters. *Environ Sci Technol* 41:3514–3520
52. Robert FC, Rominder PSS, Fu HX (2007) Free synthetic and natural estrogen hormones in influent and effluent of three municipal wastewater treatment plants. *Water Environ Res* 79:969–974
53. Morteani G, Moller P, Fuganti A, Paces T (2006) Input and fate of anthropogenic estrogens and gadolinium in surface water and sewage plants in the hydrological basin of Prague (Czech Republic). *Environ Geochem Health* 28:257–264
54. Fernandez MP, Ikononou MG, Buchanan I (2007) An assessment of estrogenic organic contaminant in Canadian wastewaters. *Sci Total Environ* 373:250–269
55. Liu ZH, Ito M, Kanjo Y, Yamamoto A (2009) Profile and removal of endocrine disrupting chemicals by using an ER/AR competitive ligand binding assay and chemical analyses. *J Environ Sci* 21:900–906
56. Cargouet M, Perdiz D, Asmaa MS, Sara TK, Levi Y (2004) Assessment of river contamination by estrogenic compounds in Paris area (France). *Sci Total Environ* 334:55–66
57. Peterson EW, Davis PK, Orndorff HA (2000) 17 β -estradiol as an indicator of animal waste contamination in mantled karst aquifers. *J Environ Qual* 29:826–834
58. Roefer P, Snyder S, Zegers RE, Rexing DJ, Fronk JL (2000) Endocrine-disrupting chemicals in a source water. *J Am Water Works Assoc* 92(8):52–58
59. Nghiem LD, Manis A, Soldenhoff K, Schäfer AI (2004) Estrogenic hormone removal from wastewater using NF/RO membranes. *J Membr Sci* 242:37–45

60. Wicks C, Kelley C, Peterson E (2004) Estrogen in a karstic aquifer ground water. *Ground Water* 42:384–389
61. Campbell CG, Borglin SE, Green FB, Grason A, Wozei E, Stringfellow WT (2006) Biologically directed environmental monitoring, fate, and transport of estrogenic endocrine disrupting compounds in water: a review. *Chemosphere* 65:1265–1280
62. Braga O, Smythe GA, Schäfer AI, Feitz AJ (2005) Fate of steroid estrogens in Australian inland and coastal wastewater treatment plants. *Environ Sci Technol* 39:3351–3359
63. Shen JH, Gutendorf B, Vahl HH, Shen L, Westendorf J (2001) Toxicological profile of pollutants in surface water from an area in Taihu Lake, Yangtze Delta. *Toxicology* 166:71–78
64. Xiao XY, McCalley DV, McEvoy J (2001) Analysis of estrogens in river water and effluents using solid-phase extraction and gas chromatography–negative chemical ionisation mass spectrometry of the pentafluorobenzoyl derivatives. *J Chromatogr A* 923:195–204
65. Drewes JE, Hemming J, Ladenburger SJ, Schaaue J, Sonzogni W (2005) An assessment of endocrine disrupting activity changes during wastewater treatment through the use of bioassays and chemical measurements. *Water Environ Res* 77:12–23
66. Rodríguez-Mozas S, López de Alda MJ, Barceló D (2004) Monitoring of estrogens, pesticides and bisphenol A in natural waters and drinking water treatment plants by solid-phase extraction–liquid chromatography mass spectrometry. *J Chromatogr A* 1045:85–92
67. Beck IC, Bruhn R, Gandrass J, Ruck W (2005) Liquid chromatography–tandem mass spectrometry analysis of estrogenic compounds in coastal surface water of the Baltic Sea. *J Chromatogr A* 1090:98–106
68. Joss A, Andersen H, Ternes T, Rihle PR, Siegrist H (2004) Removal of estrogens in municipal wastewater treatment under aerobic and anaerobic conditions: consequences for plant optimization. *Environ Sci Technol* 38:3047–3055
69. Johnson AC, Williams RJ, Simpson P, Kanda R (2007) What difference might sewage treatment performance make to endocrine disruption in rivers? *Environ Pollut* 147:194–202
70. Liu ZH, Kanjo Y, Mizutani S (2009) Removal mechanisms for endocrine disrupting compounds (EDCs) in wastewater treatment – physical means, biodegradation, and chemical advanced oxidation: a review. *Sci Total Environ* 407:731–748
71. Körner W, Hanf V, Schuller W, Kempter C, Metzger J, Hagenmaier H (1999) Development of a sensitive E-screen assay for quantitative analysis of estrogenic activity in municipal sewage effluents. *Sci Total Environ* 225:33–48
72. Isidori M, Cangiano M, Palermo FA, Parrella A (2010) E-screen and vitellogenin assay for the detection of the estrogenic activity of alkylphenols and trace elements. *Comp Biochem Physiol C Toxicol Pharmacol* 152:51–56
73. Johnson AC, Sumpter JP (2001) Removal of endocrine disrupting chemicals in activated sludge treatment works. *Environ Sci Technol* 35:4697–4703
74. Johnson AC, Belfroid A, Di Corcia A (2000) Estimating steroid estrogen input into activated sludge treatment works and observation on their removal from the effluent. *Sci Total Environ* 256:163–173
75. Ternes TA, Stumpf M, Mueller J, Haberer K, Wilken R-D, Servos M (1999) Behavior and occurrence of estrogens in municipal sewage treatment plants–I. Investigations in Germany, Canada and Brazil. *Sci Total Environ* 225:81–90
76. Andersen H, Segrist H, Hlling-Sorensen B, Trnes TA (2003) Fate of estrogens in a municipal sewage treatment plant. *Environ Sci Technol* 37:4021–4026
77. Ternes TA, Kreckel P, Mueller J (1999) Behaviour and occurrence of estrogens in municipal sewage treatment plants–II. Aerobic batch experiments with activated sludge. *Sci Total Environ* 225:91–99
78. Carballa M, Omil F, Lema JM, Llompart M, Garcia-Jares C, Rodriguez I, Gomez M, Ternes T (2004) Behaviour of pharmaceuticals, cosmetics and hormones in a sewage treatment plant. *Water Res* 38:2918–2926
79. Clara M, Strenn B, Gans O, Martinez E, Kreuzinger N, Kroiss H (2005) Removal of selected pharmaceuticals, fragrances and endocrine disrupting compounds in membrane bioreactor and conventional wastewater treatment plants. *Water Res* 39:4797–4807
80. Panter GH, Thompson RS, Beresford N, Sumpter JP (1999) Transformation of a non-oestrogenic steroid metabolite to an oestrogenically active substance by minimal bacterial activity. *Chemosphere* 38:3579–3596
81. Semião AJC, Schäfer AI (2010) Xenobiotics removal by membrane technology: an overview. In: Fatta-Kassinos D, Bester K, Kümmerer K (eds) *Environmental pollution*, vol 16. Springer, Netherlands, pp 307–338
82. Ben Aim RM, Semmens MJ (2002) Membrane bioreactors for wastewater treatment and reuse: a success story. *Water Sci Technol* 47(1):1–5
83. Manem J, Sanderson R (1996) Membrane bioreactors. In: Malleval J, Odendaal PE, Wiesner MR (eds) *Water treatment membrane processes*. McGraw-Hill, New York, pp 17.1–17.31
84. Galil NI, Sheidorf Ch, Stahl N, Tenenbaum A, Levinsky Y (2003) Membrane bioreactors for final treatment of wastewater. *Water Sci Technol* 48(8):103–110
85. Cirja M, Ivashechkin P, Schäfer A, Corvini PFX (2008) Factors affecting the removal of organic micropollutants from wastewater in conventional treatment plants (CTP) and membrane bioreactors (MBR). *Rev Environ Sci Biotechnol* 7:61–78
86. Abegglen C, Joss A, McArdell CS, Fink G, Schlüsener MP, Ternes TA, Siegrist H (2009) The fate of selected micropollutants in a single-house MBR. *Water Res* 43:2036–2046
87. Rosenberger S, Kruger U, Witzig R, Manz W, Szwedzyk U, Kraume M (2002) Performance of a bioreactor with submerged membranes for aerobic treatment of municipal waste water. *Water Res* 36:413–420

88. Weiss S, Reemtsma T (2008) Membrane bioreactors for municipal wastewater treatment – a viable option to reduce the amount of polar pollutants discharged into surface waters? *Water Res* 42:3837–3847
89. Massé A, Spérantio M, Cabassud C (2006) Comparison of sludge characteristics and performance of a submerged membrane bioreactor and an activated sludge process at high solids retention time. *Water Res* 40:2405–2415
90. Bouhabila EH, Ben Aïm R, Buisson H (1998) Microfiltration of activated sludge using submerged membrane with air bubbling (application to wastewater treatment). *Desalination* 118:315–322
91. Bouhabila EH, Ben Aïm R, Buisson H (2001) Fouling characterisation in membrane bioreactors. *Sep Purif Technol* 22:123–132
92. Zhang B, Yamamoto K, Ohgaki S, Kamiko N (1997) Floc size distribution and bacterial activities in membrane separation activated sludge processes for small-scale wastewater treatment/reclamation. *Water Sci Technol* 35(6):37–44
93. Huang X, Gui P, Qian Y (2001) Effect of sludge retention time on microbial behaviour in a submerged membrane bioreactor. *Proc Biochem* 36:1001–1006
94. Ng HY, Hermanowicz SW (2005) Membrane bioreactor operation at short solids retention times: performance and biomass characteristics. *Water Res* 39:981–992
95. Gao M, Yang M, Li H, Wang Y, Pan F (2004) Nitrification and sludge characteristics in a submerged membrane bioreactor on synthetic inorganic wastewater. *Desalination* 170:177–185
96. Cicek N, Macomer J, Davel J, Suidan MT, Audic J, Genestet P (2001) Effect of solids retention time on the performances and biological characteristics of a membrane bioreactor. *Water Sci Technol* 43(11):43–50
97. Pollice A, Giordano C, Laera G, Saturno D, Mininni G (2006) Rheology of sludge in a complete retention membrane bioreactor. *Environ Technol* 27:723–732
98. Hu JY, Chen X, Tao G, Kekred K (2007) Fate of endocrine disrupting compounds in membrane bioreactors systems. *Environ Sci Technol* 41:4097–4102
99. Spring AJ, Bagley DM, Andrews RC, Lemanik S, Yang P (2007) Removal of endocrine disrupting compounds using a membrane bioreactor and disinfection. *J Environ Eng Sci* 6:131–137
100. Urase T, Kagawa C, Kikuta T (2005) Factors affecting removal of pharmaceutical substances and estrogens in membrane separation bioreactors. *Desalination* 178:107–113
101. Clara M, Strenn B, Ausserleitner M, Kreuzinger N (2004) Comparison of the behaviour of selected micropollutants in a membrane bioreactor and a conventional wastewater treatment plant. *Water Sci Technol* 50(5):29–36
102. Kreuzinger N, Clara M, Strenn B, Kroiss H (2004) Relevance of the sludge retention time (SRT) as design criteria for wastewater treatment plants for the removal of endocrine disruptors and pharmaceuticals from wastewater. *Water Sci Technol* 50(5):149–156
103. Schröder H.Fr. (2006) The elimination of the endocrine disruptors 4-nonylphenol and bisphenol A during wastewater treatment – comparison of conventional and membrane assisted biological wastewater treatment followed by an ozone treatment. *Water Pract Technol* 1(3):0060(1–9)
104. Chen J, Hwang X, Lee D (2008) Bisphenol A removal by a membrane bioreactor. *Process Biochem* 43:451–456
105. González S, Petrovic M, Barceló D (2007) Removal of a broad range of surfactants from municipal wastewater – Comparison between membrane bioreactor and conventional activated sludge treatment. *Chemosphere* 67:335–343
106. Zuehlke S, Duennbier U, Lesjean B, Gnirss R, Buisson H (2006) Long-term comparison of trace organics removal performances between conventional and membrane activated sludge processes. *Water Environ Res* 78:2480–2486
107. Wintgens T, Gallenkemper M, Melin T (2002) Endocrine disrupting removal from wastewater using membrane bioreactor and nanofiltration technology. *Desalination* 146:387–391
108. Cirja M, Zühlke S, Ivashechkin P, Schäffer A, Corvini PFX (2006) Fate of a 14C-labeled nonylphenol isomer in a laboratory-scale membrane bioreactor. *Environ Sci Technol* 40:6131–6136
109. Dray J, Tillier F, Dray F, Ullmann A (1972) Hydrolysis of urine metabolites of different steroid hormones by β -glucuronidase from *Escherichia coli*. *Ann Inst Pasteur* 123:853–857
110. Lesjean B, Gnirss R, Buisson H, Keller S, Tazi-Pain A, Luck F (2005) Outcomes of a 2-year investigation on enhanced biological nutrients removal and trace organics elimination in membrane bioreactor (MBR). *Water Sci Technol* 52(10–11):453–460
111. Vader JS, van Ginkel CG, Sperling FM, de Jong J, de Boer W, de Graaf JS, van der Most M, Stokman PG (2000) Degradation of ethinyl estradiol by nitrifying activated sludge. *Chemosphere* 41:1239–1243
112. De Gusseme B, Pycke B, Hennebel T, Marcoen A, Vlaeminck SE, Noppe H, Boon N, Verstraete W (2009) Biological removal of 17 α -ethinylestradiol by a nitrifier enrichment culture in a membrane bioreactor. *Water Res* 43:2493–2503
113. Koh YK, Chiu TY, Boobis AR, Scrimshaw MD, Bagnall JP, Soares A, Pollard S, Cartmell E, Lester JN (2009) Influence of operating parameters on the biodegradation of steroid estrogens and nonylphenolic compounds during biological wastewater treatment processes. *Environ Sci Technol* 43:6646–6654
114. Weber S, Leuschner P, Kämpfer P, Dott W, Hollender J (2005) Degradation of estradiol and ethinyl estradiol by activated sludge and by a defined mixed culture. *Appl Microb Biotechnol* 67:106–112
115. González S, Petrovic M, Barceló D (2008) Evaluation of two pilot scale membrane bioreactors for the elimination of selected surfactants from municipal wastewaters. *J Hydro* 356:46–55
116. Andersen HR, Andersen AM, Arnold SF (1999) Comparison of short-term estrogenicity tests for identification of hormone-disrupting chemicals. *Environ Health Perspect* 107:89–108

117. Ternes TA, Andersen H, Gilberg D, Bonerz M (2002) Determination of estrogens in sludge and sediments by liquid extraction and GC/MS/MS. *Anal Chem* 74:3498–3504
118. Larsen TA, Lienert J, Joss A, Siegrist H (2004) How to avoid pharmaceuticals in the aquatic environment. *J Biotechnol* 113:295–304
119. Lai KM, Johnson KL, Scrimshaw MD, Lester JN (2000) Binding of waterborne steroid phases in river and estuarine systems. *Environ Sci Technol* 34:3890–3894
120. Yamamoto H, Liljestrand HM, Shimizu Y, Morita M (2003) Effects of physical-chemical characteristics on the sorption of selected endocrine disruptors by dissolved organic matter surrogates. *Environ Sci Technol* 37:2646–2657
121. Boxall ABA, Blackwell P, Cavallo R, Kay P, Tolls J (2002) The sorption and transport of a sulphonamide antibiotic in soil systems. *Toxicol Lett* 131:19–28
122. Clara M, Strenn B, Saracevic E, Kreuzinger N (2004) Adsorption of bisphenol-A, 17 β -estradiol and 17 α -ethinylestradiol to sewage sludge. *Chemosphere* 56:843–851
123. Mikkelsen LH (2003) Applications and limitations of the colloid titration method for measuring activated sludge surface charges. *Water Res* 37:2458–2466
124. Ganaye VA, Keiding K, Vogel TM, Viriot ML, Block JC (1997) Evaluation of soil organic matter polarity by pyrene fluorescence spectrum variations. *Environ Sci Technol* 31:2701–2706
125. Andersen HR, Hansen M, Kjolholt J, Stuer-Lauridsen F, Ternes T, Halling-Sorensen B (2005) Assessment of the importance of sorption for steroid estrogens removal during activated sludge treatment. *Chemosphere* 61:139–146
126. Ternes T, Joss A, Siegrist H (2004) Scrutinizing pharmaceuticals and personal care products in wastewater treatment. *Environ Sci Technol* 38:392–399
127. Chen X, Hu J (2009) Degradation of 17 β -estradiol and its conjugates: Effects of initial concentration and MLSS concentration. *Process Biochem* 44:1330–1334
128. Urase T, Kikuta T (2005) Separate estimation of adsorption and degradation of pharmaceutical substances and estrogens in the activated sludge process. *Water Res* 39:1289–1300
129. Joss A, Zabczynski S, Göbel A, Hoffmann B, Löffler D, McArdell CS, Ternes TA, Thomsen A, Siegrist H (2006) Biological degradation of pharmaceuticals in municipal wastewater treatment: Proposing a classification scheme. *Water Res* 40:1686–1696
130. Suárez S, Carballa M, Omil F, Lema JM (2008) How are pharmaceutical and personal care products (PPCPs) removed from urban wastewaters? *Rev Environ Sci Biotechnol* 7:125–138
131. Ying GG, Kookana RS, Kumar A, Mortimer M (2009) Occurrence and implications of estrogens and xenoestrogens in sewage effluents and receiving waters from South East Queensland. *Sci Total Environ* 407:5147–5155
132. Cicek N, Franco J, Suidan M, Urbain V, Manem J (1999) Characterization and comparison of a membrane bioreactor and a conventional activated sludge system in the treatment of wastewater containing high molecular weight compounds. *Water Environ Res* 71:64–70
133. Joss A, Keller E, Alder AC, Goebel A, McArdell CS, Ternes T, Siegrist H (2005) Removal of pharmaceuticals and fragrances in biological wastewater treatment. *Water Res* 39:3139–3152
134. Göbel A, McArdell CS, Joss A, Siegrist H, Giger W (2007) Fate of sulfonamides, macrolides, and trimethoprim in different wastewater treatment technologies. *Sci Total Environ* 372:361–371
135. Baek SH, Pagilla K (2009) Microbial community structures in conventional activated sludge system and membrane bioreactor (MBR). *Biotechnol Bioprocess Eng* 14:848–853
136. Fan XJ, Urbain V, Qian Y, Manem J, Ng WJ, Ong SL (2000) Nitrification in a membrane bioreactor (MBR) for wastewater treatment. *Water Sci Technol* 42(3–4):289–294
137. Shi J, Fujisawa S, Nakai S, Hosomi M (2004) Biodegradation of natural and synthetic estrogens by nitrifying activated sludge and ammonia-oxidizing bacterium *Nitrosomonas europaea*. *Water Res* 38:2323–2330
138. Yi T, Harper WF (2007) The link between nitrification and biotransformation of 17 alpha-ethinylestradiol. *Environ Sci Technol* 41(12):4311–4316
139. Gaulke LS, Strand SE, Kalhorn TF, David Stensel H (2009) Estrogen biodegradation kinetics and estrogenic activity reduction for two biological wastewater treatment methods. *Environ Sci Technol* 43:7111–7116
140. Wisniewski C, Leon Cruz A, Grasmick A (1999) Kinetics of organic carbon removal by a mixed culture in a membrane bioreactor. *Biochem Eng J* 3:61–69
141. Song KG, Choung YK, Ahn KH, Yun H (2003) Performance of membrane bioreactor systems with sludge ozonation process for minimization of excess sludge production. *Desalination* 157:353–359
142. Radjenović J, Matošić M, Mijatović I, Petrović M, Barceló D (2008) Membrane bioreactor (MBR) as an advanced wastewater treatment technology. *Hdb Env Chem* 55:37–101
143. Han SS, Bae TH, Jang GG, Tak TM (2005) Influence of sludge retention time on membrane fouling and bioactivities in membrane bioreactor system. *Process Biochem* 40:2393–2400
144. Sipma J, Osuna B, Collado N, Monclús H, Ferrero G, Comas J, Rodriguez-Roda I (2010) Comparison of removal of pharmaceuticals in MBR and activated sludge systems. *Desalination* 250:653–659
145. Li F, Yuasa A, Obara A, Mathews AP (2005) Aerobic batch degradation of 17- β estradiol (E2) by activated sludge: effects of spiking E2 concentrations, MLVSS and temperatures. *Water Res* 39:2065–2075
146. Yoshimoto T, Nagai F, Fujimoto J, Watanabe K, Mizukoshi H, Makino T, Kimura K, Saino H, Sawada H, Omura H (2004) Degradation of estrogens by *Rhodococcus zopfii* and *Rhodococcus equi* isolates from activated sludge in wastewater treatment plants. *Appl Environ Microbiol* 70:5283–5289

Waterborne Diseases of the Ocean, Enteric Viruses

JACQUELINA W. WOODS

US Food and Drug Administration, Center for Food Safety and Applied Nutrition, Gulf Coast Seafood Lab, Dauphin Island, AL, USA

Article Outline

Glossary

Definition of the Subject

Introduction

Enteric Viruses Found in the Environment

Detection of Enteric Viruses

Future Directions

Bibliography

Glossary

Attenuate To reduce or weaken a pathogen.

Cultivate The ability of viruses to replicate under ideal conditions.

Enteric virus Viral particle associated with human feces. Enteric viruses range in size of 20–80 nm and contain RNA or DNA enclosed by a protein capsid.

Environmental stability (microorganism) The ability of an organism to withstand degradation when exposed to environmental factors, such as temperature and sunlight.

Epidemiology The study of the causes, distribution, and patterns of health and illness in a population.

Gastroenteritis Inflammation of the stomach and large and small intestines.

Genogroup Related viruses within a genus.

Icosahedral Having a geometric structure that contains 20 identical equilateral triangular faces, 30 edges, and 12 vertices.

Open reading frame (ORF) A DNA sequence that does not contain a stop codon. ORFs are presumptive genes.

Serotype Distinct antigenic variations within a subgroup or subspecies of bacteria or viruses.

Viremia The presence of viruses in the blood spread through primary and secondary transmission.

Definition of the Subject

Human *enteric viruses* pose a significant health threat in the aquatic environment since they are transmitted via the fecal-oral route. Human activities such as faulty septic systems, agricultural runoff, urban runoff, sewage outfall, and wastewater discharge from vessels are ways enteric viruses are introduced into the environment. There are approximately 140 enteric viruses found in humans, and approximately one billion per gram of feces where at least 10% of the population can shed these viruses at any given time [1]. Enteric viruses can be transferred throughout the environment by attaching to particulates in groundwater, estuarine water, seawater and rivers, estuaries, shellfish grown in contaminated waters, and by aerosols emitted from sewage treatment plants [2]. The fate of these enteric viruses can take many routes, such as rivers, lakes, sewage, land runoff, estuaries, and groundwater. Humans can be exposed to enteric viruses through various routes: crops grown in land irrigated with wastewater or fertilized with sewage, shellfish grown in contaminated water, sewage-polluted recreational waters, and contaminated drinking water. In a waterborne disease outbreak study between 1946 and 1980, water system deficiencies that contributed to these outbreaks were categorized under five major headings: (1) use of contaminated untreated surface water, (2) use of contaminated untreated groundwater, (3) inadequate or interrupted treatment, (4) distribution network problems, and (5) miscellaneous [3]. Deficiencies in treatment and distribution of water contributed to more than 80% of the outbreaks.

Introduction

The most commonly studied enteric viruses belong to the families of single-stranded RNA viruses (ssRNA) [*Picornaviridae* (enteroviruses, polioviruses, coxsackieviruses, hepatitis A virus, and echoviruses), *Caliciviridae* (noroviruses, caliciviruses, and astroviruses)], double-stranded DNA (dsDNA) [*Adenoviridae* (adenoviruses)], and double-stranded RNA (dsRNA) [*Reoviridae* (reoviruses and rotaviruses)]. These enteric viruses have cellular and molecular structures that make them resistant to current water treatment processes. Emerging enteric viral pathogens like Aichi virus (ssRNA), sapovirus (ssRNA), and

picobirnaviridae (bi-segmented dsDNA) have properties similar to currently studied enteric viral pathogens in that they are non-enveloped, resistant to heat inactivation, stable at low pH, resistant to chlorination, or resistant to UV light inactivation [1]. Parvoviruses (the smallest known enteric viruses with ssRNA and high heat resistance) and polyomaviruses (include JC virus, BK virus, and simian virus 40 which are non-enveloped dsDNA viruses) can also be considered emerging viruses but do not cause acute *gastroenteritis* as do the most commonly studied enteric viruses [4–6].

Although enteric virus infections are associated primarily with self-limiting gastroenteritis in humans, they may also cause respiratory infections, conjunctivitis, hepatitis, and diseases that have high mortality rates, such as aseptic meningitis, encephalitis, and paralysis in immunocompromised individuals [7]. In addition, some enteric viruses have been linked to chronic diseases, such as myocarditis and insulin-dependent diabetes [1].

Human enteric viruses can be transmitted by water, food, fomites, and by human contact. They typically have a low infectious dose which makes them an immediate public health concern. In some instances, e.g., norovirus infections, the infectious dose can be as little as one to ten virions with a secondary attack rate of 50% [7]. The risk for infection when consuming viral contaminated water is at least tenfold greater than that for pathogenic bacteria with similar exposures [8]. The World Health Organization (WHO) estimates that the most common waterborne disease, diarrhea, is responsible for 4.6 billion episodes of diarrhea and causes 2.2 million deaths worldwide [9]. Enteric viruses in water are of particular concern because of the potential for contamination from a variety of sources. Because significant advances have been made in the area of environmental virology, enteric viruses have now been recognized as the causative agents in many nonbacterial gastroenteritis cases and outbreaks identified in the past as of unknown etiological origin [2]. Enteric viruses have been detected and linked to many outbreaks from contaminated waters and foodstuff [10–12].

Enteric Viruses Found in the Environment

Norovirus

An outbreak of acute gastroenteritis occurred among students and teachers in a school in Norwalk, Ohio, in

1968 [13]. The initial attack rate had a morbidity of 50%, and a secondary attack produced a higher rate of 82% [14]. Nausea and vomiting occurred in >90% of those affected, while diarrhea occurred in 38% of those affected. The clinical onset of the illness was typically 12–24 h with duration of 12–60 h. Upon subsequent transmission of stool filtrates to human volunteers in 1972, a small round-structured virus (SRSVs) ~27-nm was identified by electron microscopy [15]. Later studies revealed that other SRSVs morphologically similar to NoV caused gastroenteritis, but Norwalk virus remained the prototype of these fecal viruses [14]. Before the discovery of NoV, most cases of gastroenteritis not attributed to bacteria were thought to be caused by nonbacterial gastroenteritis [16]. In 1981, Greenberg et al. published data proposing that NoV might be a calicivirus. In 1993, Jiang et al. provided molecular evidence that NoV was a calicivirus with a 7.6-kilobase (kb) viral genome consisting of a positive sense, single-stranded, polyadenylated RNA. The genome consists of three *open reading frames* (ORFs), which code for the nonstructural proteins including the RNA polymerase (ORF1), capsid protein (ORF2), and a minor structural protein (ORF3). Noroviruses can be separated into five *genogroups* (GI, GII, GIII, GIV, and GV) based on the sequence comparison of the RNA polymerase and capsid regions [17, 18]. Genogroup I infects humans only, genogroup II infects humans and swine, and genogroup IV infects humans and canine. Genogroup III infects bovine animals, and genogroups V infects mice. Of the five genogroups, the classification scheme for the different clusters and strains identifies NoV GI containing 8 clusters, GII containing 17 clusters, and GIV containing 1 cluster (Table 1). Noroviruses have an assigned nomenclature where strains are named after the geographic location of the outbreak from which they are first described. The genogroups and genotypes were characterized and classified based on the RNA polymerase region and the complete capsid gene sequences [17, 18]. To date, the most common genogroup implicated in gastrointestinal infections is genogroup II.

Noroviruses are transmitted primarily through the fecal-oral route, consumption of fecally contaminated food or water, or by person-to-person contact. There is a high rate of secondary infection, and this can occur by airborne transmission. Outbreaks commonly occur in

Waterborne Diseases of the Ocean, Enteric Viruses. Table 1 Genogroups, clusters, and strains of human norovirus

GI		GII		GIV	
Cluster	Strain	Cluster	Strain	Cluster	Strain
4	ChibaJPN00	15	J23-USA02	1	Alpha-NLD99
5	MusgroveGBR00	4	Bristol-GB93		
2	SOV-GBR93	12	Wortley-GB00		
6	Hesse-DEU98	1	Hawaii-US94		
1	NY-USA93	16	Tiffin-US03		
3	DSV-USA93	5	Hillingd-GB00		
7	Wnchest-GBR00	2	Msham-GB95		
8	Boxer-USA02	10	Erfurt-DEU01		
		13	Faytvil-US02		
		17	CSE1-US03		
		6	Seacrof-GB00		
		8	Amstdam-NLD99		
		9	VAbeach-US01		
		14	M7-US03		
		7	Leeds-GB00		
		3	Toronto-CAN93		
		11	SW918-JPN01		

schools, nursing homes, hospitals, camps, daycare centers, cruise ships, and any other close community situation. Because norovirus is not listed as a reportable disease to health officials, estimates of the level of infection may not be accurate.

To date, the ability to *cultivate* norovirus has been futile, although several attempts have been made. A Duizer et al. study [19] utilized 27 different cell lines in an unsuccessful attempt to cultivate norovirus. Straub et al. [20] utilized human embryonic intestinal epithelial cells (INT-407) with 3-D tissue culture. The use of INT-407 cells with 3-D tissue culture appeared to be promising as this method provided the closest attempt at mimicking the structure of in vivo cells. The difficulty with cultivation of norovirus may be explained by its specific requirements or receptors needed for attachment to cells in order for replication to occur. Experiments with recombinant norovirus particles and human gastrointestinal biopsies showed preferential binding to epithelial cells of the pyloric

region of the stomach and to enterocytes on duodenal villi [19]. Human-specific blood group antigen H1-type expression was shown to be necessary for norovirus attachment to the cells, and most routine cell cultures lack the characteristics of these specialized human intestinal epithelial cells.

The majority of background information on the biological properties of norovirus has been obtained through humans who volunteer for human feeding studies [21, 22]. Infectivity can only be assessed in human dose-response experiments, and the infectious dose had been determined to be around ten virus genomes [22]. This is very critical when considering norovirus survival. Norovirus remains infectious under refrigeration and freezing conditions, it survives well in the environment, and it is resistant to heat and drying conditions. This was demonstrated in an outbreak at a long-term care facility where norovirus survived on fomites and continued to cause infection 2 weeks after the initial peak of illnesses [23]. Norovirus will

continue to be a significant health threat worldwide as this virus continues to evolve. The lack of a tissue culture cell line for effective propagation of norovirus will hinder complete understanding of how this virus causes infection in addition to decreased progression on the development of a productive vaccine.

Hepatitis A Virus

Hepatitis A is a non-enveloped RNA virus 27–32 nm in diameter. It has an *icosahedral* symmetry and belongs to the genus *Hepatovirus* of the *Picornaviridae* family. HAV has a positive-polarity single-stranded 7.5-kb genome with a single ORF with three distinct regions (P1, P2, and P3) [24]. Region P1 consists of four capsid proteins VP1–VP4. Region P2 consists of nonstructural proteins 2A–2C, and region P3 consists of nonstructural proteins and virus-specific proteins (VPg) [24]. HAV demonstrates a high degree of antigenic (amino acid) and genetic (nucleotide) conservation throughout the genome [24, 25]. Although this high propensity for conservation exists, there is still enough diversity to define HAV genotypes and subgenogroups [26]. The genomic regions commonly used to define HAV genogroups include (1) the C terminus of the VP3 region, (2) the N terminus of the VP1 region, (3) the 168-bp junction of the VP1/P2A regions, (4) the 390-bp region of the VP1/P2B regions, and (5) the entire VP1 region [26, 27]. A total of six genotypes have been identified: genotype 1A, 1B, II, III, IV, V, and VI. Genotypes I, II, and III are of human origin, and IV, V, and VI genotypes are of nonhuman primate origin. Genotype I and III are the most prevalent genotypes isolated from humans [28]. Because there is only one *serotype* of HAV, individuals infected by HAV in one part of the world are protected from reinfection by HAV in another part of the world.

Infections with HAV can produce effects that range in severity from asymptomatic to death from fulminant hepatitis. Infections with HAV are typically self-limiting and do not result in chronic liver disease. The virus shed in the feces and peak fecal excretion, hence infectivity, occurs prior to the onset of symptoms [29]. Clinical manifestations can increase with age, and with older children and adults, symptoms are typically present with jaundice occurring in 70% of those infected [29]. An average incubation period is 28 days with

a range of 15–50 days. Symptoms include gastroenteritis, fever, malaise, anorexia, nausea, abdominal discomfort, dark urine, and jaundice, all which may last up to 2 months. Chronic liver disease has not been shown to persist, although 15–20% of those infected may have prolonged or relapsing disease lasting up to 6 months [30]. Fulminant hepatitis is a rare complication in HAV infections, only occurring in <1% of those infected, with the highest rates occurring in young children and elderly adults who may have underlying liver conditions [28]. Among those with fulminant disease, reported findings demonstrate nucleotide and/or amino acid substitutions in the 5' untranslated region, P2 region, and the P3 region of the HAV genome [28, 31].

Cell culture propagation of human and nonhuman HAV has occurred in African green monkey kidney and fetal rhesus monkey kidney cells [32, 33]. Propagation of HAV of human origin is quite different than propagation of other picornaviruses of human origin. HAV of human origin requires an extensive adaptation period or serial passages before it grows in cell culture, and once it has adapted, HAV becomes *attenuated* as demonstrated by not producing disease in experimentally inoculated nonhuman primates [34]. Mutations causing attenuation in viral nucleic acid could play a significant role in the adaptation of HAV in cell culture [24, 33]. The attenuated strain HM-175 was adapted in cell culture, and this strain is currently used as vaccination agent for HAV [24].

HAV is a major cause of acute hepatitis in developed countries, while in developing countries, it can be considered endemic [35]. HAV is transmitted primarily by the fecal-oral route, via person-to-person contact, and by contaminated food or water, while the other common hepatitis viruses (hepatitis B and hepatitis C) are typically transmitted by blood or body fluids. HAV is stable in the environment when associated with organic material, and it is resistant to low pH and heating [36]. The name “hepatitis” is derived from the fact that HAV replicates in the liver and affects the liver. The source of the infectious agent, however, cannot be identified in approximately 50% of reported hepatitis A cases in the U.S., partly because of the long incubation before the appearance of symptoms [28]. Because outbreaks of HAV can cause considerable morbidity and even mortality, it is imperative that HAV be isolated and identified as the implicated pathogen.

Enterovirus

Human enteroviruses are members of Picornaviridae family and *Enterovirus* genus. They are icosahedral, non-enveloped, with a diameter of 27–30 nm. The genome is 7.5–8.5 kb and is composed of single-stranded positive polarity RNA. Enteroviruses are further divided into the subgenera/species of poliovirus, coxsackieviruses, echoviruses, and enteroviruses, and they are marked according to their serotypes. The poliovirus group consists of three different serotypes. Types 1 and 3 are recognized as epidemic, while type 2 is endemic [37]. Coxsackieviruses consist of groups A and B, where the A group contains 24 serotypes, and the B group contains six serotypes. In 2003, the International Committee on Taxonomy of Viruses classified *Enteroviruses* into four groups of species based on their molecular properties [38] (Table 2).

Infections due to enteroviruses are common, causing a range of disease including pharyngitis and poliomyelitis. In the United States, it is estimated that 30–50 million enterovirus infections occur a year, of which only 5–15 million are symptomatic [37]. The virus is spread by the fecal-oral route and person to person through direct contact with secretions of an infected individual. The incubation period is usually 3–7 days with virus transmission lasting 3–10 days after symptoms appear. Replication occurs in the gastrointestinal track, but can occur in other tissues, such as nerve and

muscle [37, 39]. Polioviruses typically infect their host by attacking the central nervous system, causing paralysis (poliomyelitis) in infected individuals. The spread of poliovirus has been limited by the development and use of vaccines. The Sabin trivalent oral live attenuated vaccine (OPV) consists of three live attenuated strains of 1, 2, and 3 serotypes grown in cell culture. The Salk vaccine is a trivalent inactivated polio vaccine (IPV) given by injection. The Sabin vaccine has an advantage over the Salk vaccine in that it elicits secretory IgA antibody production in addition to IgA, IgM, and IgG serum antibody production [40]. Coxsackieviruses have been associated with respiratory infections, gastroenteritis, insulin-dependent diabetes, myocarditis, and pericarditis [1]. Echoviruses are typically less infectious and are usually associated with the common cold and other respiratory diseases. The numbered enteroviruses have not been widely studied, but they are generally associated with bronchiolitis, conjunctivitis, meningitis, and paralysis [7, 41]. Also, enteroviruses are one of few enteric viruses which produce *viremia* in infected patients.

Enteroviruses are resistant to most concentrations of chlorine used in sewage treatment, and they are tolerant to cold and warm temperatures. This makes them ideally suited for survival in the environment. Stability of enteroviruses in the environment is dependent on temperature, humidity, and UV radiation. In order to inactivate 90% of poliovirus in salt water environment, 671 days at 4°C is required; on the other hand, an increase in the temperature to 25°C reduces the inactivation time by 25 days [37]. Symonds et al. [81] collected sewage influent and effluent from several wastewater treatment plants (WWTP) across the United States and revealed that 75% of the sewage influent and 8.3% of the effluent contained enteroviruses. In 2003, the largest European outbreak of enterovirus-related infection occurred in Belarus [42]. Over 1,300 people became ill, and water contaminated with echovirus and coxsackievirus was identified as the source of the infection. Aside from water and sewage samples, enteroviruses have been detected in food samples. In 1914, the first described food-borne outbreak was linked to milk contaminated with poliovirus [43]. After pasteurization of milk was adopted, transmission of enterovirus by contaminated milk decreased dramatically. Despite its demonstrated presence in the

Waterborne Diseases of the Ocean, Enteric Viruses.

Classification of human enteroviruses

Groups of enteroviruses	Species of enteroviruses
Enterovirus-A	Coxsackievirus A2–8, 10, 12, 14, 16
	Enterovirus 71, 76, 89, 90, 91
Enterovirus-B	Coxsackievirus A9
	Coxsackievirus B1–6
	Echovirus 1–7, 9, 11–21, 24–27, 29–33
	Enterovirus 69, 73–75, 77–78, 79–88, 100–101
Enterovirus-C	Coxsackievirus A1, 11, 13, 17–22, 24
	Poliovirus 1–3
Enterovirus-D	Enterovirus 68, 70

environment and sewage, there have been very few food-borne-related outbreaks due to enterovirus.

Adenovirus

Adenoviruses are members of the *Adenoviridae* family and the *Mastadenovirus* genus, which comprises five genera and infects hosts across the extended spectrum of vertebrates [44, 45]. Human adenoviruses are a double-stranded DNA virus containing a non-enveloped icosahedral shell with fingerlike projections forming each of the 12 vertices [46]. Its DNA is linear, about 35 kb in size, and it encodes for more than 30 structural and nonstructural proteins [47]. In 1953, the first adenovirus was isolated from human adenoid tissue [48]. There are 51 serotypes of adenovirus, and they are divided into six species, A–F, based on their hemagglutination properties, their oncogenic potential in rodents, and DNA homology or GC content [44].

Infections with adenoviruses can result in a wide range of clinical symptoms. Subgroups A, D, and F are sites for gastrointestinal infections. Subgroup B is responsible for lung and urinary tract infections. Subgroups C and E are responsible for respiratory tract infections. Serotypes 40 and 41 are the cause of most adenovirus associated–gastroenteritis, and serotypes 4 and 7 are associated with most cases of ARD (acute respiratory disease) in the United States [49]. Most adenovirus infections are self-limiting except cases where the infected individual is immunocompromised. However, in 2007, there were cases of a new emerging strain of adenovirus 14 that caused fatal respiratory disease in healthy individuals [50]. After the primary infection, immunity is conferred to the causative adenovirus serotype.

Human adenoviruses are specific to humans even though adenoviruses infect a range of animals. In 2005, Cox et al. reported no viable human adenovirus detection in feces of cattle, valve, pig, sheep, horse, dog, poultry, wombat, cat, kangaroo, possum, wood duck, rat, wild pig, fox, rabbit, ferry cat, goat, carp, and deer. In domestic sewage worldwide, human adenoviruses have been detected in high concentrations and their detection in sewage seems to have little seasonal variability [2, 44, 51]. As with most enteric viruses, adenovirus survives better in the environment and sewage treatment than the current indicator bacteria.

Adenoviruses have increased resistance to UV light, and this increased resistance could be due to the DNA repair mechanism of the host cell. Because of the *environmental stability* of adenovirus, they have been suggested as an indicator of viral pollution [52]. In current literature, adenovirus has been associated with waterborne outbreaks. Foodborne outbreaks have been suspected but not confirmed [49, 53].

Rotavirus

Rotaviruses are a significant cause of infantile viral gastroenteritis, with 114 million annual cases of diarrhea and up to 352,000–592,000 deaths in children younger than five worldwide [54]. Rotavirus' increased activity occurs during peak respiratory virus activity, rising in the winter and peaking in the spring.

Rotaviruses are members of the family *Reoviridae* and are double-stranded RNA with 11 segments and a triple capsid structure. Classification of rotavirus into P and G genotypes/serotypes are based on the outer capsid proteins, VP4 (P protein) and VP7 (G protein). Only 14 G and 15 P genotypes have been identified in humans [55]. There are seven different serogroups of rotavirus (A–G). Although individuals may have multiple rotavirus infections over time, group A rotaviruses cause the greatest numbers of cases of acute gastroenteritis (AGE) in children under 2.

The primary route of transmission of rotavirus is the fecal-oral route; additionally, the persistence of rotaviruses in aquatic environments and their resistance to water treatment and disinfection processes may facilitate their transmission to humans. Rotavirus has been implicated in many waterborne epidemics, and these viruses have been isolated in sewage, river water, groundwater, and drinking water [56–58]. Rotaviruses are resistant to most disinfectants, heat, proteolytic enzymes, and pH with ranges of 3–10. Their ability to survive in various environmental conditions for long periods of time contributes to their high rate of infection.

The burden of rotavirus disease has prompted a need for an effective vaccine. In 1998, FDA-approved RotaShield, a tetravalent human-rhesus genetic reassortant vaccine. RotaShield was withdrawn from the US market after it was shown to be associated with intussusceptions in infants. Vaccine development

continued with the license of a monovalent human rotavirus vaccine Rotarix™ and the pentavalent bovine-human reassortant vaccine RotaTeq™. To date, neither of the later derived vaccines have been associated with intussusceptions.

Hepatitis E Virus

Hepatitis E virus infections are a major cause of acute enterically transmitted non-A, non-B (ET-NANB) hepatitis in many developing countries throughout Middle East, Asia, and Africa [59]. This disease is associated with >50% of sporadic acute hepatitis which is a cause for great public health concern [60]. HEV can be both endemic and epidemic with outbreaks generally associated with fecal contamination of drinking water. North America and Europe have generally been considered non-endemic with documented cases from individuals who had previously traveled to endemic areas. It has been implied that HEV disease may be zoonotic in endemic regions and regions where HEV disease is non-endemic [61]. Domestic animals such as cattle, sheep, and pigs all appear to be reservoirs of the virus, which may explain the presence of anti-HEV antibodies in individuals with no obvious or known contact with human strains of the virus.

Hepatitis E is a single-stranded non-enveloped positive-sense RNA virus approximately 27–30 nm in diameter with a genome of about 7.2 kb [62]. HEV is currently classified as the only member of the genus *Hepevirus* in the family *Hepeviridae* [63]. Phylogenetic analysis has identified four significant genotypes of HEV [62]. Genotype 1 (G1) was isolated from Africa and Asia, genotype 2 (G2) from Mexico, and genotype 3 (G3) and genotype 4 (G4) from the United States, China, Japan, Europe, and Vietnam [62]. Genotype 3, which differs from genotypes 1, 2, and 4 by 25.6–26.3%, 25.3–25.5%, and 23.7–24.7%, respectively, is the first animal strain of HEV to be characterized [61]. Additional HEV genotypes may be present as other new strains have not been completely sequenced and therefore cannot be prominently assigned to a genotype at present time.

Recognition of Hepatitis E as a distinctly unique agent of human disease did not occur until the 1980s [64] with the development of serological laboratory

tests. Prior to the development of these tests, clinical HEV infection was determined to be the etiological agent by process of elimination. Since the *epidemiology* of hepatitis A had been well characterized, assumptions were plausible that >90% of the adult population in most Asian and Indian countries had previous exposure to HAV and were therefore immune to HAV infection [65]. This statement was proven when serologic tests for the diagnosis of hepatitis A and hepatitis E were applied to stored clinical samples collected during a waterborne epidemic of viral hepatitis in India [65]. Among these stored samples were those of the epidemic that occurred in Delhi, India, in 1955–1956 where in a population of 1.6 million, approximately 29,300 jaundice cases occurred, with an estimated 67,700 non-icteric infections. This was the first reported outbreak of disease where ET-NANB hepatitis was determined to be the etiological agent and virtually 100% of stored serum samples from such epidemics were found to contain IgG anti-HAV but not IgM anti-HAV. This can be considered definitive evidence for the existence of a previously unrecognized hepatitis as the etiological agent in past epidemics [65]. Application of current HEV immunological tests on these epidemic samples can further strengthen the evidence for ET-NANB as the causative agent in these waterborne outbreaks occurring decades ago.

The epidemiology and virology of hepatitis E suggest that HEV is less readily transmitted than HAV [64]. In South Asia, HEV incidence has been characterized by marked seasonality, with outbreaks occurring during the rainy or monsoon seasons [65]. Outbreak investigations often reveal fecally contaminated drinking water as the source of HEV infections. It appears that rain-induced flooding allows sewage to contaminate water supplies which explains the seasonal associations. Communities with inadequate sewage disposal are prone to recurrent HEV outbreaks. Low levels of community sanitation are also strongly associated with both epidemic and endemic HEV. An example of this occurred in Indonesia where the use of river water for cooking and drinking, improper disposal of human waste, and poor personal hygiene were significantly associated with increased anti-HEV IgG seroprevalence. More recent interactions between changing human culture and HEV are seen in epidemics of hepatitis E that occur among refugee camps

under substandard living conditions in Ethiopia and Somalia [66]. These findings highlight the risk of HEV to aid workers and military installations stationed in countries where HEV is epidemic. Therefore, prevention and control of HEV in epidemic areas will entail improvements in community sanitation, management of sewage distribution, and treatment of drinking water.

Enterically transmitted Hepatitis E has a secondary attack rate around 2% and a relatively low infectivity compared to hepatitis A, which has a 10–20% secondary attack rate among household contacts. It is not clear why HEV has a lower secondary attack rate than HAV. The lowered environmental stability of HEV may contribute to its lower secondary attack rate [63]. Also, studies of HEV have shown an association between the severity of the disease and inoculum size, where a larger infectious dose of HEV may be required to cause overt disease. There has been no recognizable difference in attack rates or infection rates by sex where there was documented exposure to HEV [65]. However, some studies have shown that adult men in some instances may have a greater risk of developing clinical illness. This could be due to infectious dose since men could have higher exposure rates than women. In addition, societal restrictions placed on women in some cultures may minimize their exposure to HEV [64].

Transmission of hepatitis E from mother to child has been reported with associated morbidity and mortality. A small study was conducted among ten pregnant women who were infected with HEV in their third trimester. Of those who had successful delivery, six transmitted HEV to their child. HEV antigen was found in five of the six infants, and anti-IgM was found in three of six infants [65]. Other studies have also shown a twofold increased risk of clinical HEV disease among pregnant women compared with non-pregnant women of reproductive age [65]. Both epidemic and clinical studies have shown that pregnant women have an increased likelihood of developing acute hepatitis and even fulminant hepatic failure after HEV infection. Since hepatitis infections involve the liver, it is reasonable to assume that the increased complications of pregnancy due to hepatitis can lead to increased adverse outcomes with risks of neonatal icterus. These reported findings of vertical transmission pose difficulty in accessing the risk of fecal-oral infections in newborns.

Despite our discovery of the hepatitis E, this virus continues to pose a global public health problem. There have been many advances in understanding the clinical, virological, and epidemiological characteristics of HEV, but there are still many unanswered questions. Development of an effective vaccine is currently not possible because a cell culture system that efficiently grows the virus has not been discovered, although there have been different vaccine candidates purposed. As future research of HEV continues, the epidemiologic features of this important virus will be clarified further and a protective vaccine will be developed. Perhaps then we can resolve the reasons for increased morbidity and mortality in pregnant women and determine factors which lead the persistence of protective antibodies. Once these questions have been answered, we will have better control and understanding of hepatitis E.

Detection of Enteric Viruses

In the past, detection of enteric viruses in environmental samples largely depended on whether the agent grew in cell culture. For those types that do, such as enteroviruses and adenoviruses, detection by virus replication in cell culture demonstrated infectivity as well as their presence. The ability to detect viruses by cell culture is a clear advantage when assessing whether environmental samples or foods are microbiologically hazardous. The capacity to do quantitative assays is also a bonus of cell culture. Cytopathic effect or virus-specific killing or lysing of cells is visible by ordinary light microscopy. Cell culture assays were the most widely used protocol for detection of enteric viruses until the 1990s [53, 67, 68]. While cell culture offers quantitative analysis and infectivity, the high cost, long turnaround time, and labor-intensive efforts are drawbacks. There are many cell lines suitable for growing enteric viruses. Buffalo green monkey (BGM) cell line has been shown to give higher plaques forming units per milliliter (PFU/ml) and faster CPE for coxsackieviruses and polioviruses [69].

There are a host of enteric viruses that are capable of growth utilizing cell culture, but two significant enteric viruses, human noroviruses and wild-type HAV, are difficult or unable to propagate in cell culture. Cell culture would apparently not be a proactive monitoring protocol given the length of time required to complete most assays and difficulties of propagation.

However, cell culture used in conjunction with detection emerging assays can be used to address the issue of viability while comparing the sensitivity of each representative assay.

Detection of enteric viral pathogens with molecular-based PCR assays has successfully been utilized for a number of years [17, 70–72]. Conventional PCR methodology utilizes a pair of oligonucleotides or primers, each hybridizing to one strand of double-stranded DNA (dsDNA) target. The types of primers used can be (1) random primers – short single-stranded DNA fragments with all possible combinations of bases, (2) polythymine primers – 16-base-long thymine primers that will hybridize with the polyadenine end of the mRNA, and (3) specific primers – only the targeted region specific to the primers will be amplified. The primers act as a substrate for DNA polymerase which creates a complementary strand by the way of sequential addition of deoxynucleotides. The process of PCR can be summarized into three steps: (1) denaturation – dsDNA is separated by an increase in temperature, (2) annealing – the temperature is decreased to allow the primers to anneal to the separated DNA, and (3) extension – the extension of the DNA fragment with the primers attached by addition of deoxynucleotides. For RNA viruses such as enteroviruses, NoV, and HAV, RT-PCR, or the conversion of RNA to cDNA, is necessary. During reverse transcription, a primer is required for the reverse transcriptase (RNA-dependent DNA polymerase) to initiate the synthesis of cDNA from the viral RNA. For DNA viruses, reverse transcription is not a necessary step. The final PCR product is analyzed by electrophoresis in the presence of ethidium bromide in which the correct size of the product can be examined visually by ultraviolet light. Hybridization with digoxigenin-labeled probes or genetic sequencing can be used to further identify PCR products. If the products examined are from environmental samples, cloning of the PCR products may also be required as there can be multiple strains of individual viruses present.

Real-time quantitative PCR or qPCR is used to quantitatively determine the amount of original target present in the sample [73, 74]. During a qPCR assay, the amplicon produced during each cycle can be quantified using SYBR Green (nonspecific attachment to dsDNA), or by using a fluorescent internal probe

(specific hybridization) [74]. For SYBR green assays, melt curve analysis is important to ensure that the desired DNA fragment is detected as different DNA fragments have different T_m (melting temperature). For fluorescent internal probes, fluorescence is measured during each cycle, and when the amount of fluorescence exceeds the background level (threshold level), the sample is scored as positive. The number of cycles required to reach the threshold level, commonly referred to as the cycle threshold value (C_t), correlates with the amount of target in the sample prior to amplification [73]. Real-time PCR is an excellent tool for detection of enteric viruses in environmental samples and has been used successfully to determine the concentrations of viral genomes [75, 76].

Multiplex PCR, which utilizes multiple primer sets within a single PCR reaction, can be used to simultaneously detect different groups of viruses. However, this multiple viral detection can be difficult to optimize because of the different annealing temperature requirements of dissimilar primer sets and because of the properties of the viral nucleic acids found between viral groups [41]. Real-time PCR (qPCR) has been quite successful for detection of multiple enteric viruses, because it can analyze each target independently in the same assay by using specific internal probes binding to different fluorochromes that the real-time PCR equipment can analyze independently [77, 78]. Furthermore, the PCR products can be of a similar size, providing better amplification efficiency. For viruses that grow poorly in cell culture, the detection by PCR integrated with cell culture (i.e., ICC-PCR) drastically reduces the time needed for detection [79]. The detection of enteroviruses in water can be between 3 and 14 days using cell culture, 5 days using integrated cell culture, and less than a day using direct real-time PCR. Because viruses are normally present in very low concentrations in environmental samples, the level of sensitivity of most PCR is advantageous for detection of low copy number. While the sensitivity of PCR is beneficial, the presence of inhibitory substances (i.e., humic acid or heavy metals) in concentrated environmental samples is of concern. Internal controls for real-time PCR have been developed to determine the presence of inhibitors in a sample and ensure that reaction conditions are optimal [80].

While there has been tremendous progress in molecular detection assays, complications remain. Even though most molecular-based assays are specific, sensitive, rapid, and cost efficient, there has been no development of a universal method or standardization. Perhaps future development of molecular-based assays that can establish infectivity will combine best of cell culture and PCR when it comes to the detection of enteric viruses in the environment.

Future Directions

Enteric viruses in the aquatic environment have been thoroughly studied over the years. Their persistence and distribution in the environment continues to cause threats to public health. Development of molecular techniques has allowed detection of these viral pathogens, especially those that cannot be readily propagated utilizing cell culture techniques. As we continue to make advances in understanding the pathogenic nature of these viruses, we can be instrumental in preventing disease outbreaks and improve public health.

Bibliography

Primary Literature

- Griffin DW, Donaldson KA, Paul JH, Rose JB (2003) Pathogenic human viruses in coastal waters. *Clin Microbiol Rev* 16:129–143
- Bosch A (1998) Human enteric viruses in the water environment: a minireview. *Int Microbiol* 1:191–196
- Lippy EC, Waltrip SL (1984) Waterborne disease outbreaks 1946–1980: a thirty-five year prospective. *J Am Water Works Assoc* 76:60–67
- Bofill-Mas S, Pina S, Girones R (2000) Documenting the epidemiologic patterns of polyomaviruses in human populations by studying their presence in urban sewage. *Appl Environ Microbiol* 66:238–245
- Brauniger S, Peters J, Borchers U, Kao M, Borchers U (2000) Further studies of thermal resistance of bovine parvovirus against moist and dry heat. *Int J Hyg Environ Health* 203:71–75
- Engelbrecht RS, Weber MJ, Salter BL, Schmidt CA (1980) Comparative inactivation of viruses by chlorine. *Appl Environ Microbiol* 40:249–256
- Kocwa-Haluch R (2001) Waterborne enteroviruses as a hazard for human health. *Pol J Environ Stud* 10:485–487
- Haas CN, Rose JB, Gerba CP, Regli R (1993) Risk assessment of viruses in drinking water. *Risk Anal* 13:545–552
- World Health Organization (2010) Guidelines for drinking-water quality. Recommendations, 3rd edn. World Health Organization, Geneva
- Beuret C, Kohler D, Baumgartner A, Luthi TM (2002) Norwalk-like virus sequences in mineral waters: one year monitoring of three brands. *Appl Environ Microbiol* 68:1925–1931
- Daniels NA, Bergmire-Sweet DA, Schwab KJ, Hendricks KA, Reddy S, Monroe SA, Fankhauser RL, Monroe SS, Atmar RL, Glass RI, Mead P (2000) A foodborne outbreak of gastroenteritis associated with Norwalk-like viruses: first molecular traceback to deli sandwiches contaminated during preparation. *J Infect Dis* 181:1467–1470
- Munnoch S, Ashbolt R, Coleman DJ, Walton N, Beers-Debble MY, Taylor R (2004) A multi-jurisdictional outbreak of hepatitis A related to a youth camp-implications for catering operations and mass gatherings. *Commun Dis Intell* 28:521–527
- Adler J, Zickl R (1969) Winter vomiting disease. *J Infect Dis* 119:668–673
- Atmar R, Estes MK (2001) Diagnosis of noncultivable gastroenteritis viruses, the human caliciviruses. *Clin Microbiol Rev* 14:15–37
- Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM (1972) Visualization by immune electron microscopy of a 27-nm particle associated with acute gastroenteritis. *J Virol* 10:1075–1081
- Ripley SR (1994) Infectious disease associated with molluscan shellfish consumption. *Clin Microbiol Rev* 7:419–425
- Ando T, Monron SS, Gentsch JR, Jin Q, Lewis DC, Glass RI (1995) Detection and differentiation of antigenically distinct small round structured viruses (Norwalk-like viruses) by reverse transcription-PCR and southern hybridization. *J Clin Microbiol* 33:64–71
- Zheng D, Ando T, Fankhauser RL, Beard RS, Glass RI, Monroe SS (2006) Norovirus classification and proposed strain nomenclature. *Virology* 346:312–323
- Duizer E, Schwab KJ, Neil FH, Atmar RL, Koopmans MPG, Estes MK (2004) Laboratory efforts to cultivate noroviruses. *J Gen Virol* 85:79–87
- Straub TM, Honer zu Bentrup K, Orosz-Coghlan P, Dohnalkova A, Mayer BK, Bartholomew A, Valdez CO, Bruckner-Lea CJ, Gerba C, Abbazadegan CM, Nickerson CA (2007) In vitro cell culture infectivity assay for human noroviruses. *Emerg Infect Dis* 13:396–403
- Dolin R, Blacklow NR, Dupont H, Formal S, Buscho RF, Kasel JA, Buscho RF, Kasel JA, Chames RP, Hornick R, Chanock RM (1971) Transmission of acute infectious nonbacterial gastroenteritis to volunteers by oral administration of stool filtrates. *J Infect Dis* 123:307–312
- Teunis PF, Moe CL, Liu P, Miller SE, Lindesmith L, Baric RS, Le Pendu J, Calderone RL (2008) Norwalk virus: how infectious is it? *J Med Virol* 80:1468–1476
- Wu HM, Fornek M, Schwab K, Chapin AR, Gibson K, Schwab E, Spencer C, Henning K (2005) A norovirus outbreak of a long-term-care facility: the role of environmental surface contamination. *Infect Control Hosp Epidemiol* 26: 802–810
- Cohen JL, Rosenblum B, Ticehurst JR, Daemer RJ, Feinstone SM, Purcell RH (1987) Complete nucleotide sequence of an

- attenuated hepatitis A virus: comparison with wild-type virus. *Proc Natl Acad Sci USA* 84:2497–2501
25. Lemon SM, Jansen RW, Brown EA (1992) Genetic, antigenic, and biological differences between strains of hepatitis A virus. *Vaccine* 10(suppl 1):S40–S44
 26. Robertson BH, Janson RW, Khanna B, Totsuka A, Nainan OY, Siegl G, Widell A, Margolis HS, Isomura S, Ito K, Ishizu T, Moritsugu Y, Lemon SM (1992) Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. *J Gen Virol* 73:1365–1377
 27. Hutin YJ, Pool E, Cramer EH, Nainan OV, Weth J, Williams LT, Goldstein ST, Gensheimer KF, Bell BP, Shapiro CN, Alter MJ, Margolis HS (1999) A multistate, foodborne outbreak of hepatitis A. National hepatitis A investigation team. *N Engl J Med* 340:595–602
 28. Nainan OV, Armstrong GL, Han X, Williams I, Bell BP, Margolis HS (2005) Hepatitis A molecular epidemiology in the United States, 1996–1997: sources of infection and implications of vaccination policy. *J Infect Dis* 191:957–963
 29. Lednar WM, Lemon SM, Kirkpatrick JW, Redfield RR, Fields ML, Kelly PW (1985) Frequency of illness associated with epidemic hepatitis A virus infections in adults. *Am J Epidemiol* 122: 226–233
 30. Glikson M, Galum E, Owen R, Tur-Kaspa R, Shouval D (1992) Relapsing hepatitis A: a review of 14 cases and literature survey. *Medicine (Baltimore)* 71:14–23
 31. Fujiwara K, Yokosuka O, Fukai K, Imazeki F, Saisho H, Omata M (2001) Analysis of full-length hepatitis A virus genome in sera from patients with fulminant and self-limited acute type hepatitis A. *J Hepatol* 35:112–119
 32. Flehmig B (1980) Hepatitis A-virus in cell culture I. Propagation of different hepatitis A-virus isolates in a fetal rhesus monkey kidney cell line (FRhK-4). *Med Microbiol Immunol* 168:239–248
 33. Daemer RJ, Feinstone SM, Gust LL, Purcell RH (1981) Propagation of human hepatitis A virus in African green monkey kidney cell culture: primary isolation and serial passage. *Infect Immun* 32:388–393
 34. Feinstone SM, Kapikian AZ, Purcell RH (1973) Hepatitis A detection by immune electron microscopy of a virus-like antigen associated with acute illness. *Science* 182: 1026–1028
 35. Jothikumar N, Cromeans TL, Sobsey MD, Robertson B (2005) Development and evaluation of a broadly reactive Taqman assay for rapid detection of hepatitis A virus. *Appl Environ Microbiol* 71:3359–3363
 36. Hollinger FB, Emerson SU (2001) Hepatitis A virus. In: Knipe DM, Howley PM (eds) *Fields virology*, 4th edn. Lippincott Williams & Wilkins, New York, pp 799–840
 37. Rajtar B, Majek M, Polanski L, Polz-Dacewicz M (2008) Enteroviruses in water environment – a potential threat to public health. *Ann Agric Environ Med* 15:199–203
 38. Khetsuriani N, Lamonte-Fowlkes A, Oberst S, Pallansch MA (2006) Centers for disease control and prevention: enterovirus surveillance-United States, 1970–2005. *Morb Mortal Wkly Rep* 55:1–20
 39. Colbere-Garapin F, Christodoulou C, Crainic R, Pelletier I (1989) Persistent poliovirus infection of human neuroblastoma cells. *Proc Natl Acad Sci USA* 86:7590–7594
 40. Howard RS (2005) Poliomyelitis and the postpolio syndrome. *Br Med J* 330:1314–1318
 41. Fong T, Lipp EK (2005) Enteric viruses of human and animals in aquatic environments: health risks, detection, and potential water quality assessment tools. *Microbiol Mol Biol Rev* 69: 357–371
 42. Amvrosieva TV, Paklonskaya NV, Biazruchka AA, Kazinetz ON, Bohush ZF, Fisenko EG (2006) Enteroviral infection outbreak in the Republic of Belarus: principal characteristics and phylogenetic analysis of etiological agents. *Cent Eur J Public Health* 14:67–73
 43. Jubb G (1915) The third outbreak of epidemic poliomyelitis at West Kirby. *Lancet* 1:67
 44. Jiang X, Wang M, Wang K, Estes MK (1993) Sequence and genomic organization of Norwalk virus. *Virology* 195:51–61
 45. Wigand R, Adrian T (1986) Classification and epidemiology of adenoviruses. In: Doerfler W (ed) *Adenovirus DNA*. Martinus Nijhoff, Boston
 46. Stewart PL, Fuller SD, Burnett RM (1993) Difference imaging of adenovirus – bridging the resolution gap between X-ray crystallography and electron-microscopy. *EMBO J* 12: 2589–2599
 47. Friefeld BR, Lichy J, Field J, Gronostajski RM, Guggenheimer RA, Krevolin MD, Nagata K, Hutwitz J, Horowitz MS (1984) The in-vitro replication of adenovirus DNA in the molecular biology of adenoviruses. In: Doerfler W (ed) *Current topics in microbiology and immunology*. Springer, Berlin, pp 221–225
 48. Rowe WP, Huebner RJ, Gilmore LK, Parrot RH, Ward TG (1953) Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture. *Proc Soc Exp Biol Med* 84:570–573
 49. Center for Disease Control and Prevention (2005) Adenoviruses. CDC, Atlanta [Ref Type: Pamphlet]
 50. Center for Disease Control and Prevention (2007) Acute respiratory disease associated with adenovirus serotype 14 – four states, 2006–2007. *Morb Mortal Wkly Rep* 56:1181–1184
 51. Carter MJ (2005) Enterically infecting viruses: pathogenicity, transmission and significance for food and waterborne infection. *J Appl Microbiol* 98:1354–1380
 52. Pina S, Puig M, Lucena F, Jofre J, Girones R (1998) Viral pollution in the environment and in shellfish human adenovirus detection by PCR as an index of human viruses. *Appl Environ Microbiol* 64:3376–3382
 53. Goyal SM (2006) *Viruses in food*. Springer, New York
 54. Grimwood K, Buttery JP (2007) Clinical update: rotavirus gastroenteritis and its prevention. *Lancet* 370:302–304
 55. Martinez-Laso J, Roman A, Head J, Cervera I, Rodrigues M, Rodrigues-Aval I, Picazo J (2009) Phylogeny of G9 rotavirus genotype: a possible explanation of its origin and evolution. *J Clin Virol* 44:52–57
 56. Grassi T, Bagordo F, Idolo A, Lugoli F, Gabutti G, De Donno A (2010) Rotavirus detection in environmental water samples by

- tangential flow ultrafiltration and RT-nested PCR. *Environ Monit Assess* 164:199–205
57. Abad FX, Pinto RM, Bosch A (1994) Survival of enteric viruses on environmental fomites. *Appl Environ Microbiol* 60:3704–3710
 58. Gratacap-Cavallier B, Genoulza O, Brengel-Pese K, Soule H, Innocenti-Francillard R, Bost M, Gofit L, Zmirou D, Seigneurin M (2000) Detection of human and animal rotavirus sequences in drinking water. *Appl Environ Microbiol* 66:2690–2692
 59. Ooi W, Gawoski JM, Yarbough PO, Pankey GA (1999) Hepatitis E seroconversion in United States travelers abroad. *Am J Trop Med Hyg* 61:822–824
 60. Casares-Clemente P, Pina S, Buti M, Jardi R, Martin M, Bofill-Mass S, Girones R (2003) Hepatitis E virus epidemiology in industrialized countries. *Emerg Infect Dis* 9:448–454
 61. Engle RE, Yu C, Emerson SU, Meng XJ, Purell RH (2002) Hepatitis E virus (HEV) capsid antigens derived from viruses of human and swine origin are equally efficient for detecting anti-HEV by enzyme immunoassay. *J Clin Microbiol* 40:4576–4580
 62. Li T-C, Saito M, Ogura G, Ishibashi O, Miyamura T, Takeda N (2006) Serological evidence for hepatitis E virus infection in mongoose. *Am J Trop Med Hyg* 74:932–936
 63. Emerson SU, Purcell RH (2003) Hepatitis E virus. *Rev Med Virol* 13:145–154
 64. Purcell RH (1994) Hepatitis virus: changing patterns of human disease. *Proc Natl Acad Sci USA* 91:2401–2406
 65. Labrique AB, Thomas DL, Stoszek SK, Nelson KE (1999) Hepatitis E: an emerging infectious disease. *Epidemiol Rev* 21:162–179
 66. Isaacson M, Frean J, He J, Seriwatana J, Innis BL (2000) An outbreak of hepatitis E in northern Namibia, 1983. *Am J Trop Med Hyg* 62:619–625
 67. Farrah SR, Goyal SM, Gerba CP, Wallis C, Melnick JL (1977) Concentration of enteroviruses from estuarine water. *Appl Environ Microbiol* 33:1192–1196
 68. Rao VC, Seidel KM, Goyal SM, Metcalf TG, Melnick JL (1984) Isolation of enteroviruses from water, suspended solids, and sediments from Galveston Bay; survival of poliovirus and rotavirus adsorbed to sediments. *Appl Environ Microbiol* 48:404–409
 69. Chonmaitree T, Ford C, Sanders C, Lucia H (1985) Comparison of cell cultures for rapid isolation of enteroviruses. *J Clin Microbiol* 26:2576–2580
 70. Caro V, Guillot S, Delpeyroux F, Crainic R (2001) Molecular strategy for 'serotyping' of human enteroviruses. *J Gen Virol* 82:79–91
 71. Croci L, Medici DD, Morace G, Fiore A, Scalfaro C, Beneduce F, Toti L (1999) Detection of hepatitis A virus in shellfish by nested reverse transcription-PCR. *Int J Food Microbiol* 48:67–71
 72. Chapron CD, Ballester NA, Fontaine JH, Frades CN, Margolin AB (2000) Detection of astroviruses, enteroviruses, and adenovirus types 40 and 41 in surface waters collected and evaluated by the information collection rule and an integrated cell culture-nested PCR procedure. *Appl Environ Microbiol* 66:2520–2525
 73. Gibson UE, Heid CA, Williams PM (1996) A novel method for real time quantitative RT-PCR. *Genome Res* 6:995–1001
 74. Mackay IM, Arden KE, Nitsche A (2002) Real-time PCR in virology. *Nucleic Acids Res* 30:1292–1305
 75. Donaldson KA, Griffin DW, Paul JH (2002) Detection, quantitation of enteroviruses from surface waters and sponge tissue from the Florida Keys using real-time RT-PCR. *Water Res* 36:2505–2514
 76. Heim A, Ebnet C, Harste G, Pring-Akerblom R (2003) Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *J Med Virol* 70:228–239
 77. Kageyama T, Kojima S, Shinohara M, Uchida K, Fukushi S, Hoshino FB, Takeda N, Katayama K (2003) Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription PCR. *J Clin Microbiol* 41:1548–1557
 78. Logan C, Leary JJ, O'Sullivan N (2006) Real-time reverse transcription PCR detection of norovirus, sapovirus and astrovirus as causative agents of acute viral gastroenteritis. *J Clin Microbiol* 44:3189–3195
 79. Reynolds KA, Gerba CP, Abbaszadegan M, Pepper LL (2001) ICC/PCR detection of enteroviruses and hepatitis A in environmental samples. *Can J Microbiol* 47:153–157
 80. Burkhardt W, Woods JW, Calci KR (2005) Evaluation of wastewater treatment plants' efficiency to reduce viral loading using real-time RT-PCR. In: 105 general meeting (ASM Abstract)
 81. Symonds EM, Griffin DW, Breitbart M (2009) Eukaryotic viruses in wastewater samples from the United States. *Appl Environ Microbiol* 75:1402–1409

Books and Reviews

- Berg G (1983) *Viral pollution of the environment*. CRC Press Boca Raton, FL, New York
- Cox P, Griffith M, Angles M, Derre D, Ferguson C (2005) Concentration of pathogens and indicators in animal feces in the Sydney watershed. *Appl Environ Microbiol* 71:5929–5934
- Gerba CP, Smith JE Jr (2005) Sources of pathogenic microorganisms and their fate during land application of wastes. *J Environ Qual* 34:42–49
- Greenberg HB, Valdesuso JR, Kalica AR, Wright RG, McAuliffe VJ, Kapikian AZ, Chanock RM (1981) *Proteins of Norwalk virus*. *J Virol* 37:994–999
- Hurst CJ (2000) *Viral ecology*. Academic, California
- Jiang SC (2006) Human adenoviruses in water: occurrence and health implications: a critical review. *Environ Sci Technol* 40:7132–7140
- Koopmans M (2005) Outbreaks of viral gastroenteritis: what's new in 2004? *Curr Opin Infect Dis* 18:295–299
- Koopmans M, Duizer E (2004) Foodborne viruses: an emerging problem. *Int J Food Microbiol* 90:23–41

- Leclerc H, Schwartzbrod L, Dei-Cas E (2002) Microbial agents associated with waterborne diseases. *Crit Rev Microbiol* 28: 371–409
- Lund E (1982) Waterborne virus disease. *Ecol Dis* 1:27–35
- Rheinheimer G (1992) *Aquatic microbiology*, 4th edn. Wiley, Chichester
- Varnam AH, Evans MG (2000) *Environmental microbiology*. ASM Press, Washington, DC

Waterborne Infectious Diseases, Approaches to Control

ALAN FENWICK¹, ALBIS FRANCESCO GABRIELLI²,
MICHAEL FRENCH¹, LORENZO SAVIOLI²

¹Department of Infectious Disease Epidemiology, Schistosomiasis Control Initiative, Imperial College London, London, UK

²Department of Control of Neglected Tropical Diseases, World Health Organization, Geneva, Switzerland

Article Outline

Glossary
Definition of the Subject
Introduction
Malaria
Schistosomiasis
STH (*Ascaris*, *Trichuris*, and Hookworm Infections)
Onchocerciasis (River Blindness)
Lymphatic Filariasis
Trachoma
Guinea Worm: A Worm on Its Own
Cholera and Typhoid
Dengue and Dengue Hemorrhagic Fever (DHF)
Future Directions
Bibliography

Glossary

Neglected tropical diseases A group of poverty-promoting and disfiguring diseases disproportionately affecting the poorest populations in developing countries. They are termed neglected because they lack the visibility, research support, and funding of other, more high profile, infections,

such as malaria, HIV/AIDS, and tuberculosis. Several can be treated with inexpensive, effective treatment, while for others new drugs are needed.

Waterborne infectious diseases Those diseases that rely on, or are heavily associated with, water for at least one stage of their life cycle or transmission, or affect populations closely related to areas of water.

Definition of the Subject

Waterborne diseases represent a significant and substantial burden on human health that disproportionately affects those living in the developing world. Human populations are attracted to live near surface water, particularly in areas such as sub-Saharan Africa (SSA) where it may represent a scarce resource. The development of water resources to exploit their potential has undoubtedly been of great benefit to some human populations via, for example, an increase in the amount of land suitable for agriculture provided by irrigation schemes, or the provision of hydroelectric power following the construction of dams and reservoirs. However, as an unintentional consequence such development may have unwittingly increased the extent of human waterborne infections by increasing the areas of suitable habitat for disease vectors and/or intermediate hosts, as well as attracting human populations to congregate, thereby aiding the life cycles of many infectious organisms [1–3].

The tools to treat, control and, in some cases, eliminate and eradicate some of these diseases already exist and there are examples of considerable, but not complete, success. Redoubling of efforts against such infections is advocated. In this entry, the major waterborne diseases that impact on human health, their extent, and the current position with regard to control are examined.

Introduction

Water- and Sanitation-Related Diseases Considered (in alphabetical order)

- Cholera
- Dengue and dengue hemorrhagic fever
- Guinea worm disease (Dracunculiasis)
- Lymphatic filariasis (Elephantiasis)
- Malaria
- Onchocerciasis (River blindness)

- Schistosomiasis (Bilharzia)
- Soil-transmitted helminthiasis (STH – ascariasis, trichuriasis, and hookworm)
- Trachoma
- Typhoid and paratyphoid enteric fevers

These diseases represent a diverse taxonomic group of infections stretching from viruses to multicellular macro-parasitic helminth infections. A brief description of the main infectious organisms is presented in [Table 1](#).

While malaria has (rightly) long had strong visibility and powerful political advocacy behind its control, many of the other diseases have not traditionally benefitted from such a profile. However, in the last decade, a number of individuals advocating on behalf of each of these diseases have changed the perception of them internationally. The World Health Organization has endorsed these increases in recognition and by establishing a “Department of Neglected Tropical Diseases” within the organization, it has encouraged a growing realization that each of the hitherto

Waterborne Infectious Diseases, Approaches to Control. Table 1 Classification of major waterborne diseases affecting human health. MDA: suitable for control via mass drug administration. *STH* soil-transmitted helminthiasis, *M* million, *DALYs* disability adjusted life years

Family	Disease	Type and species	DALYs	Infected globally
Protozoa	Malaria	Cerebral (<i>Plasmodium falciparum</i>)	46.5M [12]	515M clinical cases of <i>P. falciparum</i> malaria [13]
		Others (<i>P. vivax</i> ; <i>P. ovale</i> ; <i>P. malariae</i>)		
Helminth (MDA)	Schistosomiasis (Bilharzia)	Intestinal (<i>Schistosoma mansoni</i> , <i>S. japonicum</i>)	13–15M [14]	207M [2]
		Urogenital (<i>S. haematobium</i>)		
	STH (Worms)	Roundworm (<i>Ascaris lumbricoides</i>)	10.5M [15]	807M [15]
		Whipworm (<i>Trichuris trichiura</i>)	6.4M [15]	604M [15]
		Hookworm (<i>Necator americanus</i> ; <i>Ancylostoma duodenale</i>)	22.1M [15]	576M [15]
	Onchocerciasis (River blindness)	<i>Onchocerca volvulus</i>	0.5M [16]	37M [16]
Helminth (non-MDA but eradicable)	Lymphatic filariasis (Elephantiasis)	<i>Wuchereria bancrofti</i> ; <i>Brugia malayi</i>	5.8M [17]	120M [17]
	Guinea worm	<i>Dracunculus medinensis</i>	<0.1M [18]	<1,800 [18]
Bacteria (MDA)	Trachoma	<i>Chlamydia trachomatis</i>	2.3M [19]	84M [19]
Bacteria (non-MDA)	Cholera	<i>Vibrio cholera</i>		3–5 M cases annually [20, 21]
	Typhoid and paratyphoid enteric fever	<i>Salmonella enterica</i>		27.1M [22]
Virus	Dengue and dengue hemorrhagic fever	<i>Flavivirus</i>	528,000 [23]	50M [23]

under-recognized diseases requires specific expertise, targets, and interventions in order to effectively control them. However, it has also been recognized that where overlaps in control approaches exist between infections, not least with regard to the targeted populations, and the frequency and mode of treatment, then an integrated approach to control can be employed which should be more cost-effective than the separate, individual, vertical control programs which have previously been employed. Indeed, there is a growing momentum behind such an approach [4–10]. The progress made in the control of each of these waterborne diseases to date is outlined in Table 2; alongside the aim of control (i.e., whether it be morbidity control, local elimination, global eradication [see [11] for definitions of these terms]) and the control strategies employed.

Malaria

Malaria is without doubt the world's most feared and important parasitic disease as measured by the extent of its distribution, the annual number of cases, and the death toll it exerts, particularly in infants [35–36]. It is estimated that 300–500 million cases of malaria occur each year, leading to over one million deaths [35–36], of which 75% occur in children under the age of 5 [37] (Fig. 1). The symptoms of malaria vary by parasite species and by individual human host. Fever, chills, headache, muscle aches, tiredness, nausea, and vomiting are common symptoms, but infection can also result in anemia and jaundice [25]. In the absence of effective treatment, convulsions, coma, severe anemia, and kidney failure can also occur. In the case of *Plasmodium falciparum*, the most common species of malaria in Africa, the severe cerebral form can develop which will lead to death without treatment. As a result, cerebral malaria is one of the leading causes of death in under-5-year-old children in Africa. Fortunately, many people living in areas with regular malaria transmission survive malaria and develop some protective immunity after repeated infections. Recently the field of malaria control has been challenged by Bill and Melinda Gates to begin aiming for global elimination of the infection, a target that has not been seriously attempted for decades [24, 38].

Input Photo of Malaria Effects

Four species of the protozoan genus *Plasmodium* infect man (*Plasmodium falciparum*, *P. vivax*, *P. ovale*, and *P. malariae*). All four are transmitted by the bite from an infected mosquito of the genus *Anopheles* and all four undergo similar life cycles. When a mosquito bites an infected person, it ingests blood, and takes up any malaria parasites in that blood meal. The parasite then develops in the mosquito passing through a number of stages until after between 9 and 30 days, the parasite is ready to pass back to another human. The form infective to humans (the sporozoite) is to be found in the mosquito's salivary glands from where it will be injected into the next human host when the mosquito takes future blood meals.

Once back in the human host, the sporozoites migrate to the liver where they multiply and spread into the bloodstream. The liver phase can last between 8 days and several months, depending on the malaria species. Further growth and multiplication then takes place inside red blood cells, and it is the regular destruction of these red blood cells which causes the characteristic intermittent fevers. The released parasites then invade new blood cells, and the cycle of multiplication and fever continues, with most people suffering their first fever 10 days to 4 weeks following infection. Malaria kills when the *P. falciparum* parasite invades the brain and causes cerebral malaria.

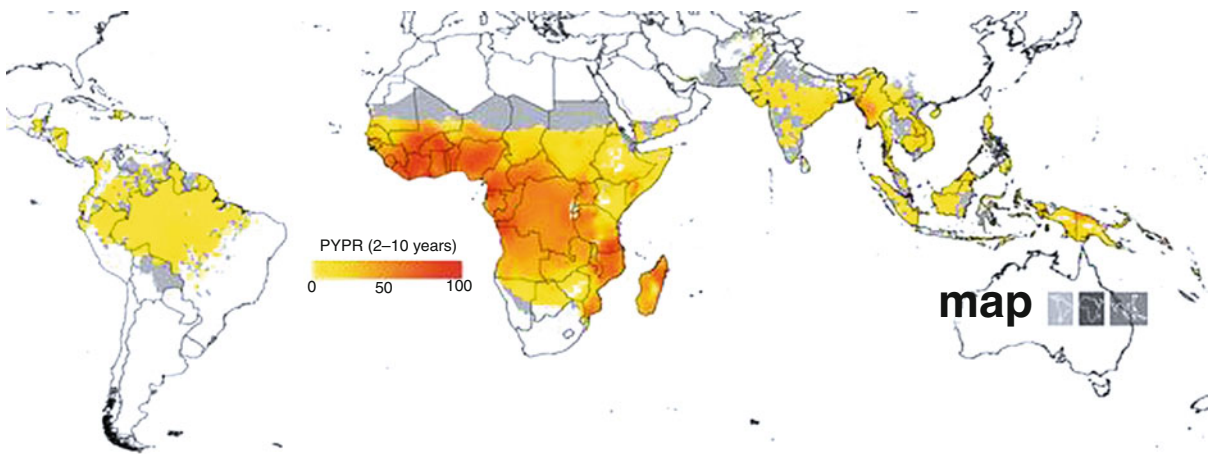
Malaria occurs mostly in tropical and subtropical countries, particularly in sub-Saharan Africa (90% of morbidity), Southeast Asia (India, Sri Lanka, Vietnam, and the Solomon Islands are endemic), and the forest fringe zones in South America (Brazil and Colombia are endemic), but its distribution is heavily dependent on the availability of suitable water bodies which support the mosquito larvae. The mosquito species vary considerably in their water and ecological requirements, in regard to levels of sunlight, aquatic vegetation, water flow, brackishness, and turbidity in a particular microhabitat. Several human factors have changed the distribution of the mosquito, for example (relatively) clean and dry urban cities may no longer support mosquitoes and therefore reduce transmission, while climate change may allow malaria to thrive in higher altitude areas than previously. The construction of irrigation systems and reservoirs, particularly in Africa, and

Waterborne Infectious Diseases, Approaches to Control. Table 2 A summary of the current status of the major waterborne infectious diseases, aims and approaches to control. ACT artemisinin combination therapy, DEC diethylcarbamazine, APOC African Program for Onchocerciasis Control, MDA mass drug administration

Disease	Current status (2011)	Control aim	Control strategy
(1) Malaria	300–500 million cases per year and up to one million deaths [13]. <i>Plasmodium falciparum</i> in Africa causes the most fatalities	Transmission control and, recently, eradication [24]	Preventing mosquito bites using long-lasting insecticide-impregnated bed nets. Presumptive treatment, speedy home-based treatment, ACTs, vector control [25]
(2a) Schistosomiasis (Bilharzia)	Over 200 million people infected; over 85% in Africa [2]. Irrigation projects and dams exacerbate the problem – treatment coverage needs to be expanded from current level of 10–20%	Morbidity control and, recently, local elimination [26]	Annual or biannual treatment with praziquantel protects children from future consequences [27]. Improved socioeconomic status, water supplies, and sanitation are necessary for elimination
(2b) Soil-transmitted helminthiasis (STH)	Over one billion of the poorest children and adults infected globally. Annual treatment coverage is now approaching 50%	Morbidity control	Six-monthly or annual deworming treatment of preschool and school-age children with albendazole or mebendazole [28]
(2c) Onchocerciasis (River Blindness)	Great progress in recent decades. 19 countries treated by APOC. Still 50 million infections in Africa because hypo-endemic areas are not treated. Elimination in some foci in Africa thought possible. Elimination being approached in Central and South America [29, 30]	Interruption of transmission (Americas)/ Morbidity control (elsewhere)	Multiple annual MDA rounds with ivermectin (Mectizan®)
(2d) Lymphatic filariasis (Elephantiasis)	120 million infected globally. A serious economic problem in Africa and the Indian subcontinent, but progress is being made toward elimination. 500 million people treated annually [31]	Global elimination as a public health problem	Multiple annual MDA rounds with albendazole and ivermectin (Mectizan®) (in Africa) or albendazole and DEC (elsewhere)
(3) Guinea worm	Rapidly approaching eradication. Fewer than 400 cases reported during the first quarter of 2011. Sudan, Ethiopia, Mali, and Chad are the countries with most residual infections (www.cartercenter.org)	Global eradication	Individual case finding and case containment, clean water provision and filtration, vector control (abate). Regular surveillance of endemic villages
(4) Trachoma (preventable blindness)	80 million infected, 8 million visually impaired [19]. Treatment expansion satisfactory: 70 million due to be treated in 2011 [32]. Has been eliminated from Morocco	Global elimination as a public health problem	Annual treatment (for 3–5 years) with azithromycin (Zithromax®), as part of a “SAFE” strategy (Surgery, Antibiotics, Face washing, and Environmental improvement) [33]
(5) Cholera	3–5 million cases and 120,000 deaths annually. A disease which follows disruption of services, clean water and sanitation after natural disasters [20]	Minimization and control of outbreaks	Up to 80% of cases can be successfully treated with oral rehydration salts. Effective control measures rely on prevention via provision of safe water and sanitation. Oral cholera vaccines can be used for control, but should not replace conventional control measures

Waterborne Infectious Diseases, Approaches to Control. Table 2 (Continued)

Disease	Current status (2011)	Control aim	Control strategy
(6) Typhoid and paratyphoid enteric fever	Conservatively estimated at 27 million infection and 215,000 deaths annually, mostly in SSA and Southeast Asia	Minimization and control of outbreaks	Improving sanitation, identification and treatment of carriers, and use of typhoid vaccines to reduce susceptibility to infection [34]
(7) Dengue and dengue hemorrhagic fever	250 million at risk and 50 million cases annually in over 100 countries, and increasing. Can be fatal [23]	Transmission and morbidity control	Effective clinical management. Replacement of fluids and sometimes blood transfusions. Vector control also required



Waterborne Infectious Diseases, Approaches to Control. Figure 1

The spatial distribution of *P. falciparum* malaria endemicity. Reproduced from work carried out by the Malaria Atlas Project [40]. *PfPR* (2–10) corresponds to the *P. falciparum* parasite rate in 2–10-year olds

changes in rice cultivation can lead to dramatic increases in the extent of areas suitable for malaria transmission. Indeed, malaria, once thought to have been eliminated in some countries, has returned as a result of the poor maintenance of water facilities and irrigation systems.

The Control of Malaria

In the 1960s, malaria was thought close to elimination after the widespread and effective use of DDT (dichlorodiphenyltrichloroethane) to control mosquito populations and chloroquine to treat infections in humans [39]. Neither of these interventions is as effective in the twenty-first century; the use of DDT has stopped due to environmental concerns and the development of resistance [39]. Equally, the parasite has

developed resistance to chloroquine, and to a succession of antimalarial drugs which followed it.

In 2011, the WHO strategy for malaria comprises four main interventions, and is implemented through the Global Malaria Program (GMP) and the Roll Back Malaria (RBM) partnership:

- Reducing mortality, particularly among children, by early case-detection and prompt treatment with effective antimalarial drugs
- Promoting the use of insecticide-treated bed nets, especially by children and pregnant women
- Prevention of malaria in pregnancy by applying intermittent preventive therapy
- Ensuring early detection and control of malaria epidemics, especially in emergency situations

In addition, vector control can be an effective control tool; mosquito breeding sites can be removed by filling in and/or draining suitable water bodies and through other environmental management schemes, while the vectors can be controlled using chemical (IRS - Indoor Residual Spraying) and biological control in certain circumstances.

Current Drugs: Artemisinin-Based Combination Therapies (ACTs)

Widespread parasite resistance has arisen to the majority of antimalarials and the management of resistance presents a key battle in the fight for malaria control and eventual eradication. The newest class of drugs, artemisinins, has a very different mode of action to conventional antimalarials, which makes them particularly useful in the treatment of resistant infections [41]. However, in order to control the development of resistance to this drug, its use is only recommended in combination with another nonartemisinin-based therapy. Artemisinin produces a very rapid amelioration of clinical symptoms and is known to cause a reduction in the transmission of gametocytes, which may help decrease the potential for the spread of resistant alleles. To date there have been very few reported side effects to drug usage, however, one great worry is that resistance to Artemisinin may develop and indeed some resistant strains may be emerging [42].

All the treatment combinations outlined below have been recommended by WHO for treatment of *P. falciparum* [43].

Artesunate and amodiaquine (*Coarsucam* and *ASAQ*). This combination has been tested and proved to be efficacious in many areas where amodiaquine retains some efficacy.

Artesunate and mefloquine (*Artequin* and *ASMQ*) has been used as an efficacious first-line treatment regimen in areas of Thailand for many years. Mefloquine adverse reactions seem to be reduced when the drug is combined with artesunate possibly due to a delayed onset of action of mefloquine. This is not likely to be an option for long-term use in Africa due to the long half-life of mefloquine, which potentially could exert a high selection pressure on parasites.

Artemether and lumefantrine (*Coartem* *Riamet*, *Faverid*, *Amatem* and *Lonart*). This combination has

been extensively tested and is effective in children under 5. It has been shown to be better tolerated than artesunate plus mefloquine combinations. There are no serious side effects documented but the drug is not recommended in pregnant or lactating women due to limited safety testing in these groups. This is the most viable option for widespread use and is available in fixed-dose formulae, which should increase compliance and adherence.

Artesunate and sulfadoxine/pyrimethamine (*Ariplus* and *Amalar plus*). This is a well-tolerated combination but the overall level of efficacy still depends on the level of resistance to sulfadoxine and pyrimethamine thus limiting its usage.

Dihydroartemisinin-piperaquine (*Duo-Cotecxin*, *Artekin*). Has been studied mainly in China, Vietnam, and other countries in Southeast Asia. The drug has been shown to be highly efficacious (greater than 90% clinical response rate).

Pyronaridine and artesunate (*Pyramax*). Manufactured by Shin Poong Pharmaceutical. Has been tested and demonstrated a clinical response rate of 100% in one trial in Hainan (an area with high levels of *P. falciparum* resistance to Pyronaridine). A multicenter phase III trial conducted in Africa found a 99.5% response rate [44].

Current Prophylactics

For travelers to malaria endemic areas who have never been exposed to malaria before, a course of prophylactics is recommended as a protection. Several options exist, as detailed in Table 3.

Waterborne Infectious Diseases, Approaches to Control. Table 3 Current prophylactic malaria tablets available for exposure-naïve individuals

Proganil	100 mg tablets are supplied as Paludrine
Chloroquine	150 mg tablets are supplied as Nivaquine or Avloclor
Mefloquine	250 mg tablets are supplied as Larium
Malarone	A combination of Atovaquone 250 mg and Proguanil 100 mg
Doxycycline	100 mg capsules or tablets

Bed Nets

Insecticide-treated nets (ITNs) were developed in the 1980s to combat malaria by preventing the mosquito from biting humans while they sleep. More recently, longer lasting insecticide nets (LLINs) have been developed which are impregnated with insecticides such as permethrin or deltamethrin. The latest production technologies (e.g., Olyset or DawaPlus) allow for a new generation of LLINs, which release insecticide for approximately 5 years. Such ITNs have been shown to be an extremely effective and cost-effective method of malaria prevention [45], aiding the drive toward Millennium Development Goal 6, which specifically calls for reduced malaria infection. These nets, which are typically available for \$2.50–\$3.50 (€2–€3) are generally purchased by donor groups like the Bill and Melinda Gates Foundation (BMGF), the President's Malaria Initiative (PMI) and the Global Fund and distributed through in-country distribution networks. One debate that has been ongoing for several years is whether or not international organizations should distribute ITNs and LLINs to people for free in order to maximize coverage and reduce price barriers. Some argue that cost-sharing between the international organization and recipients would lead to greater usage of the net because people will value them if they pay for them. Irrespective of the specifics of the approach, the main target has to be in assisting people, who most need the bed nets, to receive them and use them correctly.

Current Insecticides

Insecticides against malaria are used in two main ways: First, to kill the larvae breeding in water bodies close to human habitation; second, to kill the adult mosquitoes to prevent them biting humans inside a house. Fogging machines can be used to safely distribute insecticide in community-wide spraying in towns and tourist resorts. Unfortunately, sprayed insecticides are often expensive and the effects are only transitory, allowing mosquitoes to reinvade. There are two currently used insecticides: (1) **Malathion** – an organophosphate often used to treat crops against a wide array of insects. It can be sprayed directly onto vegetation, such as the bushes where mosquitoes like to rest, or used in a 5% solution to fog open areas. In the small amounts used for

mosquito control, it poses no threat to humans or wildlife. (2) **Permethrin** is a synthetic form of a natural insecticide (pyrethroid) which can be mixed with oil or water and applied as a mist. Both malathion and permethrin are also available in sprays for use for Indoor Residual Spraying (IRS). IRS kills adult mosquitoes that land on sprayed surfaces. Its two primary effects are to reduce the life span of vector mosquitoes so that they can no longer transmit malaria, and to reduce the density of the vector mosquitoes [46]. IRS is currently being funded by the PMI and the Global Fund. Historical and programmatic documentation has clearly established the impact of IRS [47] and expansion of coverage is likely to continue.

DDT has long been the cheapest insecticide and the one with the longest residual efficacy against malaria vectors (6–12 months depending on dosage and substrate), although environmental and resistance problems have largely curtailed its use. Other insecticides have relatively shorter residual effect (pyrethroids: 4–6 months; organophosphates and carbamates: 2–6 months).

Schistosomiasis

Three major species of schistosome worms infect humans [48]:

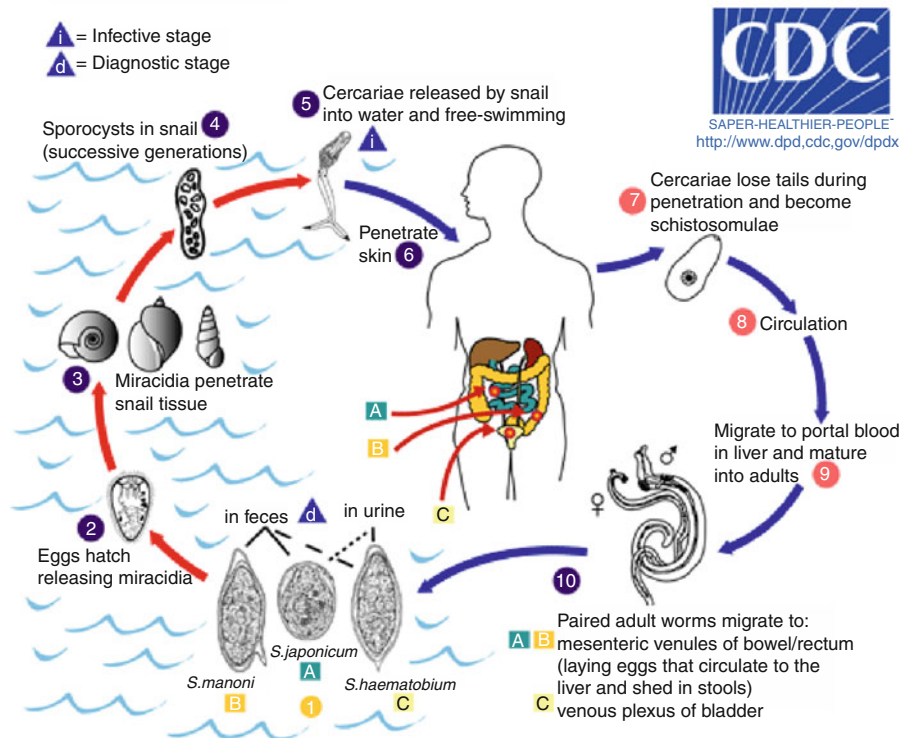
- *Schistosoma japonicum* (intestinal schistosomiasis) is found only in China and the Far East and infects humans and domestic animals. The intermediate host are amphibian snails of the genus *Onchomelania*.
- *S. mansoni* (intestinal schistosomiasis) originated in Africa, but was carried by humans to South America and the Caribbean where a suitable snail host was able to support transmission. Primates may also become infected. The intermediate hosts are snail species of the genus *Biomphalaria*.
- *S. haematobium* (urogenital schistosomiasis) is found only in Africa and the Middle East and there is no animal reservoir. The snail intermediate hosts are certain species of the genus *Bulinus*.

The worms of each species live in the blood vessels of humans and the female worm lays eggs which either escape through the urine (*S. haematobium*) or feces (*S. mansoni*/*S. japonicum*) to continue the life cycle or remain in the body trapped in the intestine wall,

bladder wall, or liver tissues. In female hosts, the worms and eggs can invade the genital system and cause damage [49]. Once the eggs have exited the body they will hatch only when they reach freshwater. The emerging free swimming larvae (miracidia) need to find and invade the correct snail intermediate host species in order to survive. Once there, they undergo asexual reproduction in their host snail over a period of a month developing through the “sporocyst” and “daughter sporocyst” intermediate stages. Eventually the next free living larval stage (cercariae) emerges from the snail into water as swimming organisms with a body and a tail. They have a short life during which time they are attracted to any human skin immersed in the water. They will attach to and penetrate unbroken skin, shedding the tail in the process. The body (now a schistosomulum) migrates to the lungs and passes through the diaphragm to reach the

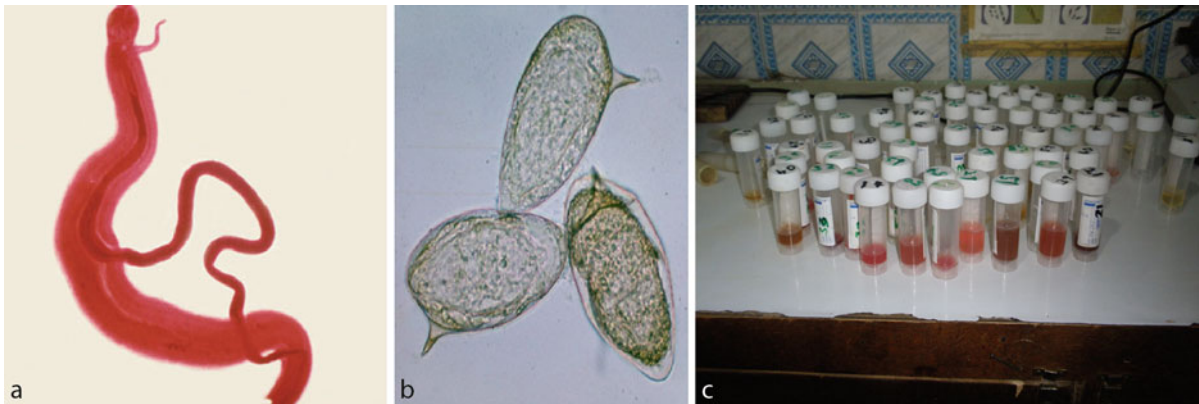
liver. About 6 weeks after initially penetrating the host, the worm in the liver will reach maturity – and about one centimeter in length. The male and female adult worms pair and the male carries the female to their selected blood vessels where they can live and produce eggs, typically for a period between 2 and 10 years [50]. *S. haematobium* worms reside around the bladder, and hence the eggs pass out in the urine. *S. japonicum* and *S. mansoni* worms reside in the blood vessels in the mesentery and hence their eggs breaking into the intestine and being voided with the feces [51–53] (Figs. 2 and 3).

An estimated 207 million people are infected globally with schistosomiasis [2] with almost 90% of those infected found in Africa [54]. The first symptom of infection is irritation at the site of penetration and is rarely noticed. A classic symptom of heavy infection is caused by the passage of schistosomula through the lungs 5–10 days after infection, but this is only seen



Waterborne Infectious Diseases, Approaches to Control. Figure 2

Life cycle of the three major schistosome species that infect humans (Taken from <http://www.dpd.cdc.gov/dpdx/html/schistosomiasis.htm>, Centers for Disease Control)



Waterborne Infectious Diseases, Approaches to Control. Figure 3

(a) A pair of schistosome adult worms, showing the female (thin worm) present in the male's gonotrophic canal, (b) Schistosome eggs, and (c) the characteristic sign of urogenital impact, blood in the urine

when many worms are passing through concurrently, as in the case of a single heavy exposure. Subsequent symptoms occur as the worms mature and start laying eggs. This can lead to blood in the urine in *S. haematobium* cases (a finding that is also used as a major diagnostic feature). *S. japonicum* and *S. mansoni* eggs breaking into the intestine can cause diarrhea and blood in the stool but this is less obvious and therefore, not a definitive diagnostic feature [48, 51–52].

Schistosomiasis is a multifaceted disease. The worms can live for up to 30 years [55, 56] and the eggs that collect in the body cause increasing organ damage over time. Heavy infections lead to severe symptoms and death from bladder cancer in *S. haematobium* infections and from portal hypertension in *S. japonicum* and *S. mansoni* infections. Uninfected people exposed to an initial heavy infection by swimming in a water body containing many infected snails can suffer extreme early symptoms, while light infections may never be detected. The norm for people residing in endemic areas is for chronic infection lasting years to occur before serious consequences are recognized. This happens because the chronic infection stage can cause significant but often undiagnosed subtle morbidity, including anemia, stunting, urinary and intestinal problems, lassitude and reduced growth, poorer school attendance and performance. Meanwhile, overt and severe organ damage can gradually occur over the course of this worsening process [48, 51, 52, 57–58] (Fig. 4).



Waterborne Infectious Diseases, Approaches to Control. Figure 4

Late stages of schistosomiasis infection, showing severe morbidity

Treatment

Currently there seems to be little chance of elimination of schistosomiasis in its transmission heartland of rural Africa because to date, elimination has only been possible in countries that have undergone significant socioeconomic development, such as Japan and Puerto Rico. Substantial improvements in water and sanitation conditions in rural African areas are required in order to permanently interrupt transmission. On a wider scale, sustained economic development is needed to eventually rid the world of this disease. Until then, the currently followed strategy is to focus on morbidity control via the large-scale administration of the drug praziquantel to at-risk groups (although recently this has been shown to also have an effect on transmission [59–60]) [61]. An annual round of treatment reaching school-age children would have a substantial effect on improving the quality of life of treated children, and protecting them from the serious consequences of their infections in later life [7].

Schistosomiasis can be treated with the generic drug praziquantel [62] (administered at 40 mg/kg body weight) which kills the adult worms. Praziquantel is currently available off-patent at a cost of about 8 USD cents (5 GBP pence) per tablet from a number of manufacturers, most of which are in India, South Korea, and China. On average 2.5 tablets are needed to treat a child and 4 tablets per adult, equating to 20 cents for the tablets needed to treat a child and 32 cents for an adult. The pharmaceutical company Merck Serono has recently committed to a drug donation program that pledges 200 million tablets over 10 years, enough to treat eight million children per year in high-prevalence areas in Cameroon, Malawi, Mozambique, Nigeria, and Senegal. It is hoped that further donations may be forthcoming in future; however, the situation is complicated as the active ingredient for the tablets is the constraining factor in the supply chain due to uncertainty in market demand. WHO is working with the drug companies to attempt to stabilize the market over the medium- to long-term, because Africa is desperately short of praziquantel ([63] and website www.schisto.org).

Improved Funding

During the 1980s several schistosomiasis control programs were started in Africa using donated drugs

(by the German developmental organization GTZ) but once the drug donations ended, the programs proved to be unsustainable and ceased. In the 1990s, three countries (Egypt, Brazil, and China) applied to the World Bank for funding and as a result they all reduced their schistosomiasis burden to below what constituted a public health problem, using a combination of drug treatment and molluscicides. The cost of molluscicides, however, proved a barrier to any expansion of their use.

At the start of this century, the Bill and Melinda Gates Foundation (BMGF) made significant funds available for treatment (from 2003 to 2006) through the Schistosomiasis Control Initiative (SCI; www.schisto.org). Six countries (Burkina Faso, Mali, Niger, Tanzania, Uganda, and Zambia) benefited from this donation and led the way in Africa for the development and implementation of national schistosomiasis programs. Thanks to the UK's Department for International Development (DFID) (2008–2013) and the United States Agency for International Development (USAID) (2006–2013), more resources have been made available (primarily through the Schistosomiasis Control Initiative [SCI] and Research Triangle International [RTI]) to expand coverage of schistosomiasis treatment throughout Africa. By working together several countries will receive the support they need to reach national coverage. Other countries plan to start implementation programs in 2011 (Côte d'Ivoire, Liberia, Malawi, and Mozambique). However, even given this expansion, many countries in need of national control programs have yet to start.

Although elimination of infection is not a target in mainland Africa, Zanzibar is aiming for elimination in 5 years from 2011; thanks to a project supported by the Gates Foundation, DFID and WHO which aims to identify the approaches (chemotherapy, water and sanitation improvement, health behaviors) needed to achieve this end point [26]. The chances of success are thought to be high because of many years of control interventions and the insular island biogeography that contributes to preventing importation of cases. Implementation will be organized by the Zanzibar government with assistance from the Ivo de Carneri Foundation on Pemba Island, The WHO, The University of Athens-Georgia, the London Natural History Museum, SCI Imperial College, and the Swiss Tropical Institute.

STH (*Ascaris*, *Trichuris*, and Hookworm Infections)

There are three separate worms which together constitute the soil-transmitted helminths (STH), so named because their life cycles do not involve an intermediate host. They are the roundworm (*Ascaris lumbricoides*), the whipworm (*Trichuris trichiura*), and the hookworm (*Necator* and *Ancylostoma* spp.). These worms inhabit the human intestine, with their eggs passed out in the feces [15–51, 52, 64].

Roundworm and whipworm cause infections associated with poor hygiene: good sanitation would prevent eggs from contaminating the environment, and regular hand washing when preparing or eating food would prevent the eggs from being ingested. Unfortunately in poor rural areas devoid of toilets, feces tend to be deposited anywhere and the roundworm and whipworm eggs are able to develop on the soil and be ready to reinfect a human host by ingestion, usually on food. Heavy infections lead to malnutrition and growth stunting in children, reduced educational achievement, and loss of future earning power [15, 65–68].

The life cycle of hookworm differs because the eggs that are passed in the feces hatch and the larvae need to pass through some developmental stages before they are ready to reinfect a human host. Infection is acquired when the free-living larvae attach themselves to the feet and ankles of passers-by. They then penetrate the skin, migrate around the body, and end up in the intestine where they attach and gorge on blood. Those individual hosts who have very heavy infections will almost certainly develop anemia as the major result, which has a serious effect on young children but also on women of childbearing age because anemia is the major cause of poor birth outcomes, and infant and maternal mortality [52, 69–71].

The nationwide elimination of these worm infections goes hand in glove with socioeconomic development, for example, these three worms used to be common in Europe and the USA, but with widespread access to safe water and sanitation, they have been eliminated from the majority of developed countries. In South Korea and Japan, infection rates were over 80% following the Second World War but during a period of 15 years this rate was reduced to less than 0.5% [72].

It is estimated that globally over a billion people in 2011 do not have the water and sanitation levels needed

to eliminate these STH and are infected with one or more worms in developing countries.

Treatment

The intestinal worms can be expelled from the body with a single 400 mg tablet of albendazole (or a 500 mg tablet of mebendazole) [62, 73]. In endemic areas where prevalence of infection is above 20%, an annual dose (or two doses when prevalence is $\geq 50\%$) of deworming tablets throughout a child's life (from the age of 1 to 14) will have an amazingly positive effect on their growth and nutritional status, their school attendance, and cognitive ability [61].

While for many years STH deworming for preschool-age children has been included into immunization interventions throughout the world [73], school-age children are usually targeted through school-based treatment interventions. A boost to STH deworming coverage has been “unintentionally” provided by the lymphatic filariasis (LF) elimination program (see LF below) because one of the drugs which has been delivered in ever increasing numbers since 2000 to eliminate LF is albendazole [74]. Approximately two billion albendazole tablets have been delivered globally since the program began [46], which means that the ancillary effect on these worms can be assumed to be significant.

Two pharmaceutical companies have pledged drug donations specifically to reduce the prevalence of STH (for a summary of these, and other drug donations, see Table 4). Johnson & Johnson has been donating 50 million tablets a year through the nongovernmental development organization (NGDO) “Children Without Worms” since 2005 and in 2010 they pledged to increase their donation of mebendazole to 200 million tablets a year, which will be mainly used in Asia and the Americas. “Feed the Children International” also donate several million mebendazole tablets through another organization, “Deworm the World,” which directs the drug to where it is needed mostly through Ministries of Education.

For Africa, GSK announced in October 2010 that from 2012 they will donate an additional 400 million albendazole tablets a year to target school-age children throughout the world. Over the next 5 years this will make a huge difference to the health and quality of life of the world's children.

Waterborne Infectious Diseases, Approaches to Control. Table 4 The contribution of the pharmaceutical industry to the control of waterborne infectious diseases

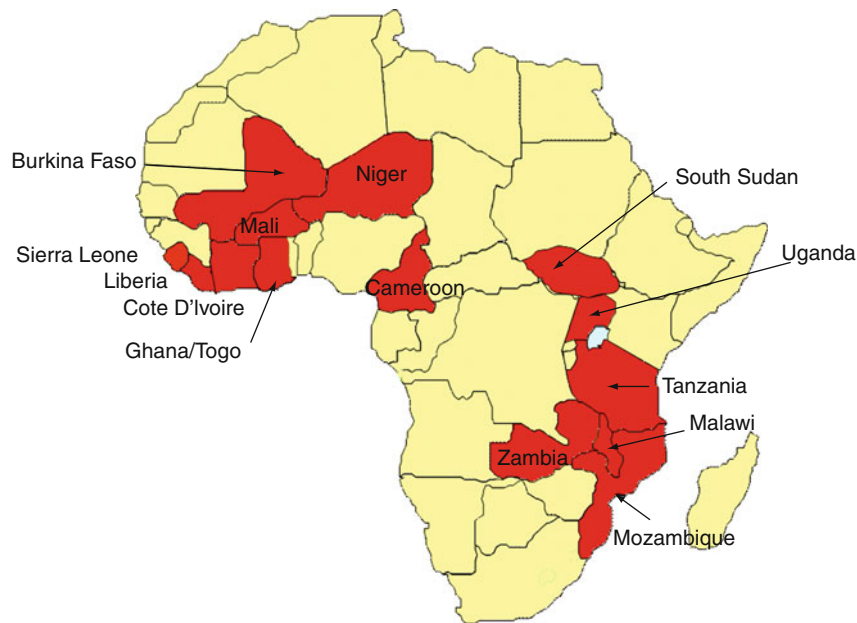
Company	Commitment
	Merck has reconfirmed their commitment (now already over 25 years) to donate ivermectin (Mectizan®) for as long as needed against both onchocerciasis and lymphatic filariasis in Africa
	GSK has already donated almost two billion tablets of albendazole against lymphatic filariasis and will continue until elimination is achieved. In October 2010, GSK committed an additional 400 million tablets a year for 5 years to deworm school-age children worldwide
	Johnson & Johnson has, for several years, donated 50 million tablets of mebendazole per year for intestinal worms – from 2012 this will be increased to 200 million tablets per year
	Pfizer committed to provide doses of azithromycin (Zithromax®) for trachoma In 2009 alone they donated 50 million tablets; in 2011 that number will reach 70 million
	In October 2010, Eisai committed to provide two billion tablets of DEC against lymphatic filariasis which is used with albendazole outside of Africa
	Medpharm (a generic manufacturer) has contributed praziquantel against schistosomiasis in the years 2004–2006
	Merck Serono has agreed to donate 200 million tablets of praziquantel over 10 years

Meanwhile, both DFID and USAID support these donations by providing essential logistical and financial support for the distribution of the tablets. DFID currently funds delivery in eight African countries (Côte d'Ivoire, Liberia, Malawi, Mozambique, Niger, Tanzania [including Zanzibar], Uganda, and Zambia), and USAID in ten countries

(see Fig. 5) and coverage is likely to expand during 2012–2015.

Onchocerciasis (River Blindness)

Onchocerciasis is a “filarial” worm that is endemic mainly in Africa and the Americas near fast-flowing rivers or streams where the host fly breeds. The adult



Waterborne Infectious Diseases, Approaches to Control. Figure 5

Countries in sub-Saharan Africa currently operating national or subnational schistosomiasis control programs

worm lives in nodules in the human body and gives birth to larvae (microfilariae) which migrate around the skin waiting to be picked up when a black fly (*Simulium* species) bites. The larvae develop in the black fly and are then transmitted back to another human host when the fly next bites [51, 52].

Prior to the 1970s this parasite was devastating to African populations. The larvae were responsible for great discomfort due to the itching they caused. However, much worse was the widespread blindness caused by the infection near the great rivers in Africa because the larvae produced by the adult worms in the human body migrate across the eye. As recently as 1970, up to 50% of adult populations living on the banks of fast moving rivers suffered impaired vision due to the persistent biting of the black fly [75–78].

The control of river blindness represents a major success in disease control, perhaps, matched only by smallpox eradication in 1979–1980 and latterly the progress toward the elimination of LF (see below). Three major programs (OCP, APOC, and OEPA), an insecticide (DDT), a pharmaceutical company (Merck), and a drug (ivermectin, Mectizan®) contributed to this success which is commemorated by a statue in front of the WHO headquarters in Geneva [79] (Fig. 6).

The OCP (Onchocerciasis Control Program) was initiated in the 1970s in West Africa to control onchocerciasis and provided DDT to be sprayed into rivers to kill the black fly larvae which breed in these fast-moving waters and therefore, to eliminate disease transmission. Nine countries implemented control as part of the OCP and as a result their endemic onchocerciasis was controlled and over a 20-year-period infection rates and blindness were reduced to very low levels [80–81].

Sadly the widespread use of DDT was shown to be environmentally unacceptable, but a major breakthrough saved the program when the drug ivermectin (Mectizan®) was found to have the effect of destroying microfilariae for a year after a single annual dose [82]. The drug manufacturer, Merck and Co., Inc., recognized that poor people in Africa would never be able to afford to purchase this drug, and so agreed to donate the drug to all those who live in endemic areas “for as long as needed” [29]. Today, some 25 years later, this donation continues, and in October 2010 the Merck commitment was renewed. Since 1986 Mectizan® has been distributed widely in OCP countries (and subsequently in 19 African Program for Onchocerciasis Control [APOC] countries) reaching over 50 million



Waterborne Infectious Diseases, Approaches to Control. Figure 6

Statue at the World Health Organization headquarters in Geneva, of a child leading an adult blinded by onchocerciasis through the village (Taken from http://www.cdc.gov/globalhealth/ntd/diseases/oncho_burden.html, Photo Credit: Carter Center)

people annually, and thereby saving millions of people from losing their sight [83].

The OCP closed down when its objectives had been met by the year 1996 but the APOC was then launched to oversee the distribution of Mectizan® in 19 additional African countries with the help of a number of NGOs (The Carter Center, Sightsavers, CBM, Helen Keller International, World Vision) [84]. The APOC mandate has recently been renewed to continue treatment until 2020, and during that time will look at the possibility of a changed strategy (doubling the dosage) to bring about elimination in some foci instead of merely control.

APOC and its partners have not managed complete coverage in Africa due to a complication to what should be a simple control program. Another parasitic filarial worm, *Loa loa*, which also produces live larvae that circulate in the body, coexists with onchocerciasis in some areas in Central Africa. Some patients with *Loa loa* when treated with Mectizan® can suffer disastrous consequences, such as encephalopathy [85]. Thus in parts of Central Africa where *Loa loa* is endemic, it is not safe to control onchocerciasis (nor LF) using MDA because of the possible dangerous complications. This is a matter yet to be resolved and further discussion is advocated [86].

The Onchocerciasis Elimination Program for the Americas (OEPA) is a regional initiative in the Americas with the goal of eliminating morbidity of river blindness and interrupting transmission in six endemic countries: Brazil, Colombia, Ecuador, Guatemala, Mexico, and Venezuela. The OEPA strategy is to encourage the endemic countries to provide sustained Mectizan® mass treatment every 6 months with the aim of reaching at least 85% of 500,000 persons estimated to be at risk of the disease [87]. All six endemic countries have established effective national programs and there are no new cases of blindness attributable to onchocerciasis in the American region. As of 2011, transmission is continuing in only two countries, Brazil and Venezuela, with surveillance in place in the remaining four [88].

In Africa, the plan for the future is continuation with the successful control program and annual Mectizan® treatments, but for the first time WHO are considering with selected governments the idea of intensifying the strategy with the aim of interrupting transmission permanently. It is thought that in many countries that have been offering annual Mectizan® for many years, the worms are beginning to die out and so by increasing frequency of treatment for a short period elimination may be achievable [89].

Lymphatic Filariasis

Two filarial worms *Brugia malayi* and *Wuchereria bancrofti* infect 120 million people globally with lymphatic filariasis (LF), with a further 600 million at risk of infection. The intermediate host of these worms are mosquitoes of several genera, and the global distribution is very extensive – Africa, the Indian subcontinent, many Pacific Islands, the Caribbean, and tropical America [90]. The deformity and misery caused by this disease are horrific. The adult worms live in and block the lymphatic system, which prevents the drainage of lymph fluid. This causes swelling of lower limbs, secondary infections, and, in the case of some men, the scrotum becomes grotesquely swollen [91]. Meanwhile, the millions of larvae which the females produce circulate in the bloodstream to be picked up by a mosquito for the transmission cycle to be completed.

The Global Alliance for the Elimination of Lymphatic Filariasis (GAELF) is supported by many partners (see http://www.filariasis.org/who_we_are/partners.html) and is housed at the Centre for Neglected Tropical Diseases (CNTD) in Liverpool. This alliance is probably the most successful disease control alliance ever formed as evidenced by the expansion of treatment coverage against this terrible affliction (see [31, 90] and website www.filariasis.org). The principle is that an annual combined dose of albendazole with Mectizan® in Africa and albendazole with DEC outside of Africa given to the whole community will effectively prevent transmission for a year. This is because the concentration of microfilariae in the blood will be so reduced that the biting mosquitoes will not pick them up [74]. Since the filarial adult worms live for an average of just 6 years elimination can be a reality if transmission can be interrupted for six consecutive years [92–94], resulting in huge health and economic benefits [95–97].

The elimination program is being underpinned by the donation pledges of pharmaceutical companies, dating back to the late 1990s. GSK made a commitment to donate albendazole, and Merck & Co. Inc. extended their onchocerciasis donation of Mectizan® to include LF in Africa [98]. Elsewhere DEC was purchased until 2010 when EISAI made their donation for the future. Currently, approximately 600 million people receive treatment annually against LF [46]. Delivery is being

funded by individual governments with assistance from the BMGF, USAID and DFID. This emphasizes the capacity of major pharmaceutical companies to work together on a major global health problem.

One unfortunate note is that many people in Africa who need treatment for LF are not yet receiving it as some governments are yet to embrace the value that the LF program provides. Approximately 80 million people are currently being treated annually with free drugs in Africa but because of the *Loa loa* complication [99] and because of the civil unrest in many parts of Africa poor coverage in some areas has not yet been resolved.

Nevertheless, the LF elimination program has achieved some fantastic results: Egypt and Zanzibar [100], and most recently Costa Rica, Suriname and Trinidad and Tobago have stopped MDA because they have successfully achieved their elimination targets [31, 90, 101]. The new donation from EISAI in Japan, amounting to two billion tablets a year, will be sufficient to treat at least the Indian subcontinent (and possibly the four countries in the American continent that are still endemic), and raising hopes of elimination of LF by 2020.

However, the drug treatments do not reverse the disfiguring morbidity that has already developed. For these cases, an individual care approach is needed based on hygiene and surgery, where appropriate (Fig. 7).

Case management of such patients has expanded and improved, and leg washing has increased the care of elephantiasis patients and has led to reductions in leg swellings by preventing secondary infections. Hydrocele surgery is slowly becoming more widespread as “surgery camps” are being established in different countries [102]. For example, in Niger, SCI supported over 600 hydrocele operations in 2010. These have been extremely successful and hundreds of men have come forward requesting the free-of-charge 20-min surgical intervention, performed under local anesthetic. The cost of each intervention at one of SCI’s camps averages less than \$200, which includes pre- and postoperative treatment (www.schisto.org).

Trachoma

Blindness due to trachoma is caused by the after effects of conjunctivitis due to *Chlamydia trachomatis* infections carried by flies. Trachoma is simply a disease of

poor hygiene, combined with dry and dusty conditions, and poverty. After severe conjunctivitis the eyelids become affected by scarring, and eyelashes are turned inward toward the eye and damage, and eventually destroy the cornea over time, leading to irreversible blindness. Improved water and sanitation would soon lead to elimination of trachoma. Meanwhile,

treatment with Zithromax[®] and regular hand and face washing is all that is needed for prevention of new cases [19, 103] (Fig. 8).

For existing trichiasis cases simple surgery can correct the problem. Face washing and improved water and sanitation are essential components in a trachoma elimination strategy. The World Health Organization



Waterborne Infectious Diseases, Approaches to Control. Figure 7

Morbidity caused by long-term lymphatic filariasis to the (a) legs, lymphedema and (b) scrotum, hydrocele



Waterborne Infectious Diseases, Approaches to Control. Figure 8

(a) Trachoma is caused by the bacteria *Chlamydia trachomatis* which is often spread by flies in areas of poor sanitation; (b) Trachoma's major effects include corneal damage and irreversible blindness

promotes the global elimination of blinding trachoma by 2020 (GET 2020) through an integrated approach named the SAFE strategy – S for surgery, A for Antibiotics, F for face washing, and E for environmental improvements to reduce the infestation of flies. The International Trachoma Initiative (ITI), established in 1999, and a number of other NGOs are very active in promoting eye care and control of trachoma, possibly combined with that of onchocerciasis where relevant. Excellent progress has been made in some countries, and ITI's flagship program has contributed to eliminating trachoma from Morocco [103].

Early treatment is effective: azithromycin – Zithromax®, an antibiotic for active infection, which is currently donated by Pfizer needs to be administered annually. In 2011, it is estimated that approximately 70 million doses of Zithromax® will be distributed. A new website www.trachoma.org has been developed to provide information about trachoma and the treatments provided by ITI. The target for reaching elimination of blinding trachoma globally remains as “2020.” As of this year:

- About 41 million people have active trachoma needing treatment.
- About 8.2 million people are estimated to have an advanced stage infection and therefore are blind or at risk of blindness.
- Fifty-seven countries are endemic, with up to 1.2 billion people living in at-risk areas.
- Since 1999, Pfizer has donated 225 million doses of Zithromax®.

Guinea Worm: A Worm on Its Own

Imagine having a worm a meter in length crawling around your body just under your skin. When the female adult worm is ready to lay her eggs, she emerges through the skin around the ankle. This subsequent burning sensation causes the victim to bathe the leg in water to cool it down. This is exactly what not to do because the worm emerges and deposits thousands of larvae into the water. These larvae infect water fleas where they develop and subsequently reinfect humans when the fleas are ingested when water is directly used for drinking. This is Guinea worm (Fig. 9).

The life-cycle of Guinea worm involves no animal host, thereby reducing the number of possible control



Waterborne Infectious Diseases, Approaches to Control. Figure 9

A photograph of the adult Guinea worm (dracunculiasis) slowly being extracted from its emergence point, here on the top of the foot (Taken from <http://www.maximsnews.com/news20081209guineawormeradication10812090803.htm>, Photo Credit: Photo © The Carter Center/Louise Gubb, Ghana, 2007)

targets. The control strategy is therefore based on identification of infected individuals, reporting and containment of cases, on behavioral change so those infected individuals do not contaminate the water, and on improved water supplies. Such combined efforts could and should lead to the eradication of Guinea worm disease. People can only get infected by swallowing infected water fleas. This happens only when people are forced to drink water from open ponds which contain these fleas, and this only happens when the water they drink has not been filtered.

Thus eradication of this unique worm was recognized to be possible some 20 years ago, and WHO and the Carter Center have been working toward this end ever since. However, there is no drug treatment for Guinea worm; the only course open to an infected person is the slow extraction of the very long worm from the skin by gently winding the worm around a stick and pulling it out over several days.

The Guinea Worm program is now moving toward eradication. The number of annual cases has been dramatically reduced from over one million in 1988 to fewer than 2,000 in 2010; and the disease remains endemic in just four countries – Chad, Ethiopia, Mali, and Sudan (www.cartercenter.org; [104]). A number of countries have been certified as free of transmission,

and there are several more countries which have not reported cases during the previous year (Burkina Faso, Cote d'Ivoire, Ghana, Kenya, Niger, Nigeria, Togo) and are considered to be in precertification phase. Once a country has been certified, there is a continued need for surveillance – in order to prevent reestablishment of transmission until global eradication is achieved.

Cholera and Typhoid

The bacterial infections cholera (caused by *Vibrio cholerae*) and typhoid (caused by *Salmonella typhi*) are endemic in countries where poor water and sanitation prevail. However, they cause their greatest damage when they reach epidemic outbreaks during times of crisis such as after natural disasters: flood, earthquakes, and tsunamis. During these periods many displaced persons congregate in temporary camps with poor sanitation, and where drinking water sources are contaminated. These really are diseases that thrive when basic services have broken down, and can quickly overwhelm public health services in such locations [105]. Cholera bacteria are most often associated with waterborne transmission, but can also be transmitted by contaminated food [105]. The classic symptoms of cholera are profuse, watery diarrhea, and vomiting, which can rapidly lead to severe dehydration and death in hours if left untreated. In addition to these severe sequelae of infection, there will likely be many, unreported cases that are asymptomatic or result in nonspecific mild symptoms [106]. Treatment of cholera via oral and intravenous rehydration therapy has resulted in reductions in fatality rates, but it still remains a dreaded disease [20]. The consequences of infection are vomiting and diarrhea, and treatment is patient care and rehydration. In epidemics about 5% of those infected will die, but if rehydration can be provided, recovery is usually achieved. Typhoid fever is caused by the bacterium *Salmonella enterica* serotype Typhi (*S. typhi*). Improvements in water and sanitation virtually eliminated typhoid from the majority of the developed world, yet it still exerts a significant public health burden in the developing world [107]. The most common mode of transport is through contaminated water, or food that has come into contact with contaminated water, and thus is heavily associated with poor water and sanitation. The burden of disease is

thought to be highest in children under the age of 5 who suffer the most serious sequelae of infection [108]. The emergence of drug resistance to a succession of antibiotics, most recently to quinolones, has put great pressure on public health systems in resource-poor settings [107].

Dengue and Dengue Hemorrhagic Fever (DHF)

These are acute febrile diseases, found in the tropics, and caused by four closely related virus serotypes, transmitted by the *Aedes* mosquito which bites during the day. The disease is now endemic in more than 100 countries in Africa, the Americas, the Eastern Mediterranean, Southeast Asia, and the Western Pacific [109]. Southeast Asia and the Western Pacific are the most seriously affected. Before 1970 only nine countries had experienced DHF epidemics, a number that had increased more than fourfold by 1995. Not only is the number of cases increasing as the disease is spreading to new areas, but explosive outbreaks are also occurring. In 2007, Venezuela reported over 80,000 cases, including more than 6,000 cases of DHF [110].

- During epidemics of dengue, infection rates among those who have not been previously exposed to the virus are often 40–50%, but can reach 80–90%.
- An estimated 500,000 people with DHF require hospitalization each year, a very large proportion of whom are children. About 2.5% of those affected die.
- Without proper treatment, DHF fatality rates can exceed 20%. Wider access to medical care from health providers with knowledge about DHF – physicians and nurses who recognize its symptoms and know how to treat its effects – can reduce death rates to less than 1%.

Early treatment with therapy to tackle shock due to hemoconcentration reductions and bleeding is crucial. Increased oral fluid intake is recommended to prevent dehydration. In extreme cases, internal gastrointestinal bleeding may occur requiring a transfusion.

Unfortunately, progress on control of this infection has been poor. Dengue fever is becoming more prevalent in some areas, which is cause for great concern. The spread of dengue is attributed to expanding geographic distribution of the four dengue viruses and their mosquito vectors, the most important of which is the

predominantly urban species *Aedes aegypti*. A rapid rise in urban mosquito populations is bringing ever greater numbers of people into contact with this vector, especially in areas that are favorable for mosquito breeding, for example, where household water storage is common and where solid waste disposal services are inadequate.

Future Directions

Great progress has been made in some areas of the battle against waterborne diseases. Control of malaria has long captured the scientific community's attention and with the recent Bill and Melinda Gates call to aim for eradication hopefully significant further progress will be accomplished. Control and elimination of some of the other waterborne diseases, such as some of the NTDs (schistosomiasis, STH, onchocerciasis, LF, trachoma), can be achieved through inexpensive (or even donated), effective, and easily administered drugs. Even more promisingly, Guinea worm is on the verge of becoming the world's second major human infectious disease to be eradicated from the globe (following smallpox [111]). This has been achieved with no available drug; but purely through health education and behavior change. For other diseases, like dengue, cholera, and typhoid, the picture is bleaker; increases in the number of cases and their distribution demand that further tools be developed and greater effort is expended in their control.

Clearly, for all these waterborne diseases discussed here, permanent change will be aided hugely by improvements in water and sanitation, particularly in the transmission heartlands of sub-Saharan Africa and Southeast Asia. Naturally, people will continue to be drawn to areas providing sources of water, and the exploitation of such resources will play a crucial role in socioeconomic development of less-developed countries. A multisectorial approach to disease control is advocated, for example, when carrying out the development of water resources (drinking water, sanitation, dams, hydroelectric power), a greater integration between engineering and health sectors is required to effectively combat these waterborne diseases permanently. In the mean time, it is advocated to use whatever tools are currently available to reduce infection and transmission and ease the burden of these diseases.

Bibliography

1. Fenwick A (2006) Waterborne infectious diseases – could they be consigned to history? *Science* 313:1077–1081
2. Steinmann P et al (2006) Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect Dis* 6:411–425
3. Keiser J, Singer BH, Utzinger J (2005) Reducing the burden of malaria in different eco-epidemiological settings with environmental management: a systematic review. *Lancet Infect Dis* 5:695–708
4. Fenwick A et al (2009) The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002–2008. *Parasitology* 136:1719–1730
5. Hotez PJ et al (2009) Rescuing the bottom billion through control of neglected tropical diseases. *Lancet* 373:1570–1575
6. Hotez PJ et al (2006) Incorporating a rapid-impact package for neglected tropical diseases with programs for HIV/AIDS, tuberculosis, and malaria. *PLoS Med* 3:e102
7. Lammie PJ, Fenwick A, Utzinger J (2006) A blueprint for success: integration of neglected tropical disease control programmes. *Trends Parasitol* 22:313–321
8. Zhang Y et al (2010) Control of neglected tropical diseases needs a long-term commitment. *BMC Med* 8:67–75
9. Kolaczinski JH et al (2007) Neglected tropical diseases in Uganda: the prospect and challenge of integrated control. *Trends Parasitol* 23:485–493
10. Linehan M et al (2011) Integrated implementation of programs targeting neglected tropical diseases through preventive chemotherapy: proving the feasibility at national scale. *Am J Trop Med Hyg* 84:5–14
11. Molyneux D, Hopkins DR, Zagaria N (2004) Disease eradication, elimination and control: the need for accurate and consistent usage. *Trends Parasitol* 20:347–351
12. WHO (2004) World health report 2004: changing history. World Health Organization, Geneva. <http://www.who.int/whr/2004/en>
13. Snow RW et al (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 435:214–217
14. King C (2010) Parasites and poverty: the case of schistosomiasis. *Acta Trop* 113:95–104
15. de Silva NR et al (2003) Soil-transmitted helminth infections: updating the global picture. *Trends Parasitol* 19: 547–551
16. Dadzie Y, Neira M, Hopkins D (2003) Final report of the conference on the eradicability of onchocerciasis. *Filaria J* 2:2
17. Zagaria N, Savioli L (2002) Elimination of lymphatic filariasis: a public health challenge. *Ann Trop Med Parasitol* 96:S3–S13
18. Centre C (2011) Dracunculiasis eradication: the final inch. <http://www.cartercenter.org/news/documents/doc2229.html>
19. Kumaresan JA, Mearns JW (2003) The global elimination of blinding trachoma: progress and promise. *Am J Trop Med Hyg* 69:S24–S28
20. WHO (2006) Cholera, 2005. *Wkly Epidemiol Rec* 81:297–308

21. WHO (2003) Cholera unveiled. Global Task Force on Cholera Control, Editor, Geneva. http://whqlibdoc.who.int/hq/2003/WHO_CDS_CPE_ZFK_2003.3.pdf
22. Crump JA, Luby SP, Mintz ED (2004) The global burden of typhoid fever. *Bull World Health Organ* 82:346–353
23. Suaya JA, Shepard DS, Beatty ME (2006) Dengue: burden of disease and cost of illness. Working paper for the scientific working group on dengue research, convened by the special programme for research and training in tropical diseases, Geneva, 1–5 October 2006
24. Roberts L, Enserink M (2007) Did they really say... eradication? *Science* 318:1544–1555
25. WHO (2008) Global malaria control and elimination: report of a technical review. World Health Organization, Geneva
26. Knopp S et al (in press) From morbidity control to transmission control: time to change tactics against helminths on Unguja Island, Zanzibar. *Acta Trop*. doi: 10.1016/j.actatropica.2011.04.010
27. Richter J (2003) The impact of chemotherapy on morbidity due to schistosomiasis. *Acta Trop* 86:161–183
28. WHO (2002) Prevention and control of schistosomiasis and soil-transmitted helminthiasis. In: World health organization technical report. World Health Organization, Geneva, pp I–v1
29. Thylefors B (2008) The Mectizan donation program (MDP). *Ann Trop Med Parasitol* 102(S1):S39–S44
30. Basáñez MG et al (2006) River blindness: a success story under threat? *PLoS Med* 3:e371
31. GAELF (2010) Half-time in LF elimination: teaming up with NTDs. In: Sixth meeting of the global alliance to eliminate lymphatic filariasis, Seoul, 1–3 June 2010
32. Mariotti SP, Pascolini D, Rose-Nussbaumer J (2009) Trachoma: global magnitude of a preventable cause of blindness. *Br J Ophthalmol* 93:563–568
33. West S (2003) Blinding trachoma: prevention with the SAFE strategy. *Am J Trop Med Hyg* 69:S18–S23
34. Crump JA, Mintz ED (2010) Global trends in typhoid and paratyphoid fever. *Clin Infect Dis* 50:241–246
35. Breman JG (2001) The ears of the hippopotamus: manifestations, determinants, and estimates of the malaria burden. *Am J Trop Med Hyg* 64S:1–11
36. Snow RW et al (1999) Estimating mortality, morbidity, and disability due to malaria among Africa's non-pregnant population. *Bull World Health Organ* 77:627–640
37. Brooker S et al (2008) Malaria in African schoolchildren: options for control. *Trans R Soc Trop Med Hyg* 102:304–305
38. Greenwood B (2009) Can malaria be eliminated? *Trans R Soc Trop Med Hyg* 103(S1):S2–S5
39. Kapp C (2000) WHO wins reprieve for DDT against malaria. *Lancet* 356:2076
40. Hay SI et al (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med* 6(3):e1000048
41. Chawira AN et al (1987) The effect of combinations of qinghaosu (artemisinin) with standard antimalarial drugs in the suppressive treatment of malaria in mice. *Trans R Soc Trop Med Hyg* 8:554–558
42. Lim P et al (2009) Pfmdr1 copy number and artemisinin derivatives combination therapy failure in falciparum malaria in Cambodia. *Malar J* 8:11
43. WHO (2006) Guidelines for the treatment of malaria. World Health Organization, Geneva
44. Tshefu AK et al (2010) Efficacy and safety of a fixed-dose oral combination of pyronaridine-artesunate compared with artemether-lumefantrine in children and adults with uncomplicated *Plasmodium falciparum* malaria: a randomised non-inferiority trial. *Lancet* 375:1457–1467
45. Maxwell CA et al (2002) Effect of community-wide use of insecticide-treated nets for 3–4 years on malarial morbidity in Tanzania. *Trop Med Int Health* 7:1003–1008
46. WHO (2011) Malaria: Indoor residual spraying. http://www.who.int/malaria/vector_control/irs/en/index.html
47. Pluess B et al (2010) Indoor residual spraying for preventing malaria (review). *The Cochrane Libr* 4:CD006657
48. Gryseels B et al (2006) Human schistosomiasis. *Lancet* 368:1106–1118
49. Poggensee G, Feldmeier H (2001) Female genital schistosomiasis: facts and hypotheses. *Acta Trop* 79:193–210
50. Fulford AJ et al (1995) A statistical approach to schistosome population dynamics and estimation of the life-span of *Schistosoma mansoni* in man. *Parasitology* 110:307–316
51. Crompton DW, Savioli L (2007) Handbook of helminthiasis for public health. Taylor and Francis, Boca Raton
52. Cook G, Zumla A (eds) (2008) Manson's tropical diseases, 22nd edn. W.B. Saunders, London
53. Mahmoud AAF (2001) Schistosomiasis. Imperial College Press, London
54. Fenwick A (2006) New initiatives against Africa's worms. *Trans R Soc Trop Med Hyg* 100:200–207
55. Warren KS et al (1974) *Schistosomiasis mansoni* in Yemeni in California: duration of infection, presence of disease, therapeutic management. *Am J Trop Med Hyg* 23:902–909
56. Harris ARC, Russell RJ, Charters AD (1984) A review of schistosomiasis in immigrants in western Australia, demonstrating the unusual longevity of *Schistosoma mansoni*. *Trans R Soc Trop Med Hyg* 78:385–388
57. de Clercq D et al (1998) The relationship between *Schistosoma haematobium* infection and school performance and attendance in Bamako, Mali. *Ann Trop Med Parasitol* 92:851–858
58. Vennerveld BJ, Dunne DW (2004) Morbidity in schistosomiasis: an update. *Curr Opin Infect Dis* 17:439–447
59. Miguel E, Kremer M (2004) Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72:159–217
60. French MD et al (2010) Observed reductions in *Schistosoma mansoni* transmission from large scale administration of praziquantel in Uganda: a mathematical modelling study. *PLoS Negl Trop Dis* 4:e897
61. WHO (2006) Preventive chemotherapy in human helminthiasis. Coordinated use of anthelmintic drugs to control interventions: a manual for health professionals and programme managers. World Health Organization, Geneva

62. WHO (2009) WHO model formulary. Based on the 15th model list of essential medicines 2007. World Health Organization, Geneva
63. Hotez P et al (2010) Africa is desperate for praziquantel. *Lancet* 376:496–498
64. Bethony J et al (2006) Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *Lancet* 367:1521–1532
65. Nokes C et al (1991) Geohelminth infections and academic assessment in Jamaican children. *Trans R Soc Trop Med Hyg* 85:272–273
66. Nokes C et al (1992) Parasitic helminth infection and cognitive function in school children. *Proc R Soc Lond B* 247:77–81
67. Simeon D et al (1994) School performance, nutritional status and trichuriasis in Jamaican schoolchildren. *Acta Paediatr* 83:1188–1193
68. Hutchinson SE et al (1997) Nutrition, anaemia, geohelminth infection and school achievement in rural Jamaican primary school children. *Eur J Clin Nutr* 51:729–735
69. Albonico M et al (2006) Intervention for the control of soil-transmitted helminthiasis in the community. *Adv Parasitol* 61:311–348
70. Brooker S, Hotez PJ, Bundy DAP (2008) Hookworm-related anaemia among pregnant women: a systematic review. *PLoS Negl Trop Dis* 2:9
71. Smith JL, Brooker S (2010) Impact of hookworm infection and deworming on anaemia in non-pregnant populations: a systematic review. *Trop Med Int Health* 15:776–7959
72. Yokogawa M (1985) JOICFP's experience in the control of ascariasis within an integrated programme. In: Crompton DW, Nesheim MC, Pawlowski ZS (eds) *Ascaris and its public health significance*. Taylor and Francis, London, pp 265–278.
73. Albonico M et al (2008) Controlling soil-transmitted helminthiasis in pre-school-age children through preventive chemotherapy. *PLoS Negl Trop Dis* 2:e126
74. Ottesen EA (2006) Lymphatic filariasis: treatment, control and elimination. *Adv Parasitol* 61:395–441
75. Kirkwood B et al (1983) Relationships between mortality, visual acuity and microfilarial load in the area of the onchocerciasis control programme. *Trans R Soc Trop Med Hyg* 77:862–868
76. Prost A (1986) The burden of blindness in adult males in the Savanna villages of West Africa exposed to onchocerciasis. *Trans R Soc Trop Med Hyg* 80:525–527
77. Murdoch ME et al (2002) Onchocerciasis: the clinical and epidemiological burden of skin disease in Africa. *Ann Trop Med Parasitol* 96:283–296
78. Little MP et al (2004) Incidence of blindness during the onchocerciasis control programme in western Africa, 1971–2002. *J Infect Dis* 189:1932–1941
79. Boatin BA, Richards FO (2006) Control of onchocerciasis. *Adv Parasitol* 61:349–394
80. Hougard J-M et al (1997) Twenty-two years of blackfly control in the onchocerciasis control programme in West Africa. *Parasitol Today* 13:425–431
81. Boatin B (2008) The onchocerciasis control programme in West Africa (OCP). *Ann Trop Med Parasitol* 102(S1): S13–S17
82. Dadzie KY et al (1991) Onchocerciasis control by large-scale ivermectin treatment. *Lancet* 337:1358–1359
83. Tielsch JM, Beeche A (2004) Impact of ivermectin on illness and disability associated with onchocerciasis. *Trop Med Int Health* 9:A45–A56
84. Amazigo U (2008) The African programme for onchocerciasis control (APOC). *Ann Trop Med Parasitol* 102(S1):19–22
85. Gardon J et al (1997) Serious reactions after mass treatment of onchocerciasis with ivermectin in an area endemic for *Loa loa* infection. *Lancet* 350:18–22
86. Padgett JJ, Jacobsen KH (2008) Loiasis: African eye worm. *Trans R Soc Trop Med Hyg* 102:983–989
87. Sauerbrey M (2008) The onchocerciasis elimination program for the Americas. *Ann Trop Med Parasitol* 102(S1):S25–S29
88. Onchocerciasis Elimination Program for the Americas (OEPA) (2011) Distribución geográfica de la oncocercosis y situación de la transmisión en las Américas. <http://www.oepa.net/epidemiologia.html>
89. Cupp EW, Sauerbrey M, Richards FO Elimination of human onchocerciasis: history of progress and current feasibility using ivermectin (Mectizan®) monotherapy. *Acta Trop* 120 S1:S100–S108
90. GAELF (2011) Global programme to eliminate lymphatic filariasis: progress report on mass drug administration, 2010. *Wkly Epidemiol Rec* 86:377–388
91. Nienga SM et al (2007) Chronic clinical manifestations related to *Wuchereria bancrofti* infection in a highly endemic area in Kenya. *Trans R Soc Trop Med Hyg* 101:439–444
92. Gyapong JO et al (2005) Treatment strategies underpinning the global programme to eliminate lymphatic filariasis. *Expert Opin Pharmacother* 6:179–200
93. Gambhir M et al (2010) Geographic and ecologic heterogeneity in elimination thresholds for the major vector-borne helminthic disease, lymphatic filariasis. *BMC Biol* 8:22
94. Michael E et al (2006) Mathematical models and lymphatic filariasis control: endpoints and optimal interventions. *Trends Parasitol* 22:226–233
95. Ottesen EA et al (2008) The global programme to eliminate lymphatic filariasis: health impact after 8 years. *PLoS Negl Trop Dis* 2:e317
96. Hooper PJ et al (2009) The global programme to eliminate lymphatic filariasis: health impact during its first 8 years (2000–2007). *Ann Trop Med Parasitol* 103(S1):S17–S21
97. Chu BK et al (2010) The economic benefits resulting from the first 8 years of the global programme to eliminate lymphatic filariasis. *PLoS Negl Trop Dis* 4:e708
98. Hopkins AD, Molyneux DH (2009) A decade of ivermectin-albendazole donation for lymphatic filariasis. *Ann Trop Med Parasitol* 103(S1):S3
99. Molyneux DH (2009) Filaria control and elimination: diagnostic, monitoring and surveillance needs. *Trans R Soc Trop Med Hyg* 103:338–341

100. Mohammed KA et al (2006) Progress towards elimination lymphatic filariasis in Zanzibar: a model programme. *Trends Parasitol* 22:340–344
101. Rawlins SC et al (2004) Evidence for the interruption of transmission of lymphatic filariasis among schoolchildren in Trinidad and Tobago. *Trans R Soc Trop Med Hyg* 98:473–477
102. Brantus P (2009) Ten years of managing the clinical manifestations and disabilities of lymphatic filariasis. *Ann Trop Med Parasitol* 103(51):S5–S10
103. Knirsch C (2007) Trachoma: ancient scourge, disease elimination, and future research. *Curr Infect Dis Rep* 9:21–28
104. Barry M (2007) The tail end of guinea worm – global eradication without a drug or a vaccine. *N Engl J Med* 356:2561–2564
105. Sack DA et al (2004) Cholera. *Lancet* 363:223–233
106. Glass RI et al (1982) Endemic cholera in rural Bangladesh. *Am J Epidemiol* 116:959–970
107. Bhutta Z (2006) Endemic cholera in rural Bangladesh, 1966–1980. *Br Med J* 333:78–82
108. Sinha A et al (1999) Typhoid fever in children aged less than 5 years. *Lancet* 354:734–737
109. WHO (1997) Dengue haemorrhagic fever: diagnosis, treatment prevention and control. World Health Organization, Geneva
110. San Martin JL et al (2010) The epidemiology of dengue in the Americas over the last three decades: a worrisome reality. *Am J Trop Med Hyg* 82:128–135
111. Henderson DA (1976) The eradication of smallpox. *Sci Am* 235:25–33

Waterborne Parasitic Diseases in Ocean

ROBIN M. OVERSTREET

Department of Coastal Sciences, Gulf Coast Research Laboratory, University of Southern Mississippi, Ocean Springs, MS, USA

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Parasites Acquired When People Eat Infected Seafood Products

Parasites Acquired When People Have Direct or Indirect Contact with Them

Parasites That Produce Excretions or Secretions That Result in an Immunological Response in People

Management, Control, and Treatment of Parasites Affecting People

Future Directions

Acknowledgments

Bibliography

Glossary

Autoinfection Reinfection of a host by the progeny of a parasite already present in the same host individual without exposure to the external environment.

Cyst/Encapsulation A cyst is a structure to aid dissemination or protection of a parasitic stage derived from the parasite and in some cases additionally from the host. A capsule differs by being derived from the host, often as a cellular response.

Erythema Abnormal redness of the skin due to local congestion as in inflammation, caused by irritation or injury to the tissue.

Hermaphrodite parasite An individual containing both male and female gametes that can be functional concurrently or with one following the other; some able to form zygotes.

Hypersensitivity reaction A damaging and sometimes fatal reaction produced by the normal immune system, which requires a pre-sensitized, immune state of the host.

Life cycle of parasite The orderly sequence of distinct stages through which the agent progresses in the course of development to maturity or sexual stage.

Life history of parasite The life cycle of an agent including facultative paratenic hosts, feedback strategies, interaction with the environment, and other ecological influences.

Paratenic host A specific host, also termed “transport host,” which acquires a stage of an agent that does not develop to the subsequent stage and can be either acquired by another paratenic host or by a final host; that host can be critical in completing a life cycle.

Pathogenesis The production or development of a disease, specifically the cellular reactions and other pathologic mechanisms occurring in the progression of the disease.

Reservoir host A definitive host that serves as an alternant relative to the host of interest for an agent that disseminates infective stages.

Salinity An expression of concentration of salts dissolved in water, including primarily chloride,

sodium, magnesium, sulfur, calcium, and potassium but with other elements in low concentrations. It is usually expressed as parts per thousand (ppt) or parts per million (ppm) and usually based on the electrical conductivity ratio of the sample to “Copenhagen water,” an artificial seawater manufactured to serve as a world “standard” or more recently in practical salinity units (psu) as the conductivity ratio of a seawater sample to a standard KCl solution. Full strength seawater is considered 35 ppt, or 35 g of salt per liter of solution; salinity in estuaries can fluctuate from 0 to over 35 ppt daily, seasonally, or yearly, depending on winds, rain, currents, temperature, and geography.

Urticaria An allergic reaction comprising pale, pink, focal swellings, or wheals, on the skin that itch, burn, or sting; also referred to as “hives.”

Zoonosis An animal disease transmissible to humans under natural conditions or a human disease transmissible to animals.

Definition of the Subject and Its Importance

Several parasites that infect marine and coastal animals or contaminate the marine and coastal waters can infect humans and present a significant public health risk. Historically, parasites from the marine environment were ignored as representing a risk. More recently, more researchers have investigated actual or potential human marine infections; easily observable worms and other metazoans have been experimentally studied to determine if they can infect model mammals; some humans are now more likely to come into contact with infective parasites; immunocompromised people are more susceptible to parasitic infections; molecular methods have been used to differentiate known or potential infective parasitic species; improved molecular and serological methods have been developed to detect known protozoan and metazoan parasites; and infectious parasites have been investigated in detail. As discussed in this chapter, some parasites involved with marine waters and seafood products can kill large numbers of people, other fatal ones are rare, many cause human illness, and infections or the possibility of infections keeps people from eating seafood or entering marine waters. Moreover, the

economic loss of seafood products from actual or perceived problems involving parasites in seafood caused by media exposure can be significant [1, 2] just as it is when involving perceived contamination of seafood by oil or other toxicants (gulfseagrant.tamu.edu/oilspill/pdfs/latourismPerception_BPOilWave1.pdf).

Introduction

Textbooks once considered parasites of marine origin or those in the marine environment as posing no public health risk. Over the last few decades, more and more parasites in different groups have been determined to be zoonotic agents that either are known to infect humans or can serve as a potential public health risk based on experimental infections in nonhuman mammalian hosts or are closely related to known human-infecting agents. These parasites include metazoans and “protozoans” that occur in water or in seafood products. Protozoa occurs in quotation marks because its historic higher taxonomic groups are not necessarily closely related [3] and some include stages with more than one cell. For example, a single myxosporidian spore may develop from a stage consisting of 13 cells, and molecular data demonstrate no relationship between Myxosporea and any group consisting of members with single cells and recognized as a protozoan. At least one metazoan taxonomic group, ascaridoid nematodes, has members that produce excretions and secretions (ES) that are potentially harmful to susceptible people.

Numerous factors are involved with risk of infection. For example, the likelihood of a visitor or traveler getting infected by eating raw seafood in well-known urban Japanese or other Asian restaurants is relatively low, but dining in rural local restaurants or street shops has a higher risk of infection with a variety of parasites [4]. The highest worldwide risk for a helminth infection is from preparation of raw or inadequately cooked seafood at home. This is usually because restaurants typically use products historically known not to be infected in the region from which they were collected or the products were frozen and then thawed. Public health risk also occurs for personal contact with parasites for people using coastal and marine water for either occupation or recreation.

An apparent increase in human infections has occurred because of increased opportunities to get

infected. (1) A global spread and increased popularity of “exotic cuisines” at least partially relates to increased human infections because of the increase in various media such as television and magazines. This has led to an increase in homemade raw seafood products, which are much more likely to contain infective parasites than products served in restaurants. (2) The increased popularity of raw products also relates to an increase and translocation of people throughout the world. People that move or visit abroad take their cuisines with them. They may eat infected products in their new surroundings that are similar to the uninfected products they ate in their native country. They can also spread their diseases as they travel, allowing new intermediate hosts and reservoirs to become established. (3) Much of the world’s population is becoming more affluent, and with that more people are able to try new cuisines. Also, there is more opportunity to partake of established cuisines. For example, in Vietnam the improved economy has increased the traditional social activities that are conducive to eating more raw seafood and, consequently, acquiring more parasites. (4) With the resulting increase in people and demand for healthy fishery products, there has been increased utilization of different fishing grounds, species, and stocks. Even though some stocks are overfished, methods are now available for ships to go farther to sea or deeper to catch replacement products, possibly with improved transportation and storage systems. When a specific fish is assumed to be free from parasites harmful to humans, when the same species is caught in a different location, when it is maintained on a vessel longer, or when it is produced on fish farms, that resulting fish product may have or it may acquire harmful parasites. Also, fishing a previously underutilized species may result in obtaining an infected stock. (5) Regulations such as those protecting specific animals may result in an increased public health risk. For example, protecting seal pups and reducing seal fashions increases the final host population for *Pseudoterranova* spp., which results in a heavily infected cod intermediate host population. Also, the need to discharge fisheries wastes from commercial and subsistence vessels results in increased infections in fishes and marine mammals that feed on the wastes. Alternatively, regulations in some countries are lacking in regard to fishery products or for

imported fishery products but not necessarily all meats, leaving consumers with the false impression that the products are safe. (6) Climatic and environmental health conditions are always changing, and, with those changes, infection dynamics of parasites quickly change. For example, during the El Niño of 1997–1998, the junction of the warmwater Kuroshio Current from the south with the cool Oyashio Current from the north moved 3,000 km northward from Kyushu to Hokkaido, Japan. This migration of the current’s junction and its temperature elevation of 3.4°C caused an increase in abundance of krill, the intermediate host of *Pseudoterranova azarasi* and members of the *Anisakis simplex* complex. The sea lion final host of *P. azarasi* occurred in Hokkaido but not Kyushu, and, in addition, protected cetaceans that are the final hosts of *A. simplex* sensu lato migrated north so as to feed heavily on krill or on fish that ate the krill, acquiring the juvenile nematodes from them. Consequently, the Japanese used to eating nearly parasite-free seafood from Hokkaido began acquiring infections. In contrast, pollution from Southeast Asia kept some fish intermediate hosts of *A. simplex* sensu lato from migrating to Japan, especially to southern Kyushu, reducing the number of anisakiasis cases there [5]. Each few years, climate in some regions worldwide changes enough to influence the seafood production and the parasitic infections. (7) Reduced resistance in some people because of disease, organ transplants, or immunosuppressive drug therapy relate to an increased susceptibility to various parasites, especially parasites that can replicate in or on humans with being challenged by an additional dose of parasites.

Increased documentation of human infections results from increased technical abilities. (1) Improved media coverage allows people to know when an epidemic occurs. It also stimulates potentially infected people to get examined, sometimes unnecessarily and to the detriment of the seafood industry. (2) Diagnostic tests are becoming better and more available both because of research plus technical advancements and because various funds are made available to survey regions or satisfy needs. (3) The same can be said for accurate identifications. Molecular means make it easy to detect parasites that cannot be seen grossly or a juvenile that cannot be morphologically associated with an adult, upon which the species has been named

and described. (4) With good identifications and detection methods, more critical examinations can be obtained and compiled, allowing for useful epidemiological assessments.

The organization involves five sections. The first treats parasites that people acquire from the marine environment by eating seafood. The second treats those where people get infected with parasites by contact with them, either directly by active penetration by the parasites or indirectly by passive association with the agents. In the third, people get disease from allergens from parasites or parasite products. This is followed by a section on management, control, and treatment of the infectious parasites. The final section briefly examines what is required in the future to reduce, eliminate, or keep infections under control. In the first section on acquisition of parasites by eating inadequately prepared food, examples are divided into those that mature in humans and those that do not. The reason for this is that worms that mature in humans are voided or produce eggs that can be detected in fecal examinations. Those parasites that do not mature in the patient or produce cyst dispersal stages are more difficult to detect without immunological or invasive techniques. Moreover, the consumer not able to provide proof of a parasitic infection is less likely to find medical help, and the medical community is less likely to specifically diagnose, report the condition, or show a follow-up interest in the case. Several of these agents that have not yet been reported are treated as potential public health risks. The first two sections are each divided by taxonomic groups. In some cases the same parasites or members of the same parasitic group fit into more than one section, and the possible confusion is discussed in the different cases.

Parasites Acquired When People Eat Infected Seafood Products

Zoonotic parasitic diseases acquired by eating inadequately prepared infected seafood products include most of those treated by disease researchers and reported in recent literature. Some of the responsible agents are seen grossly, such as 5–10 m of senile cestode strobilae protruding out of the anus of alarmed seafood consumers or a 4-cm-long ascaridoid nematode being

vomited or passing out through the nose of someone recently eating uncooked cod. Most zoonotic diseases are recognized by the signs of a disease or fortuitous findings of diagnostic features. Not all infections are diseases. A few hundred small marine heterophyids probably can line the human intestine without apparent symptoms, but more may cause pathological alterations and result in disease. A lot of factors involving the patient and the history of the infection can influence the alterations. Perhaps easier to envision would be a non-marine example of the hookworm *Necator americanus*. A barefoot boy can pick up a few dozen juveniles penetrating his feet from the warm, moist soil. There may be a slight itching and rash where the juveniles enter, but as they are carried to the lungs, pass up the respiratory tract to the mouth, are swallowed, and ultimately attach to the intestine as 7- to 11-cm-long adults, they produce no symptom. On the other hand, increase the number, say above 50, and pathological changes can occur during all the phases. Most important, once attached, the blood loss becomes enough to produce anemia and protein deficiency, and the resulting loss of iron and protein may retard growth and mental development. This will be associated with abdominal pain, diarrhea, loss of appetite and weight, and “disease.” Numbers of worms alone can shift an “infection” without disease to a “disease.”

Parasites That Mature in Humans

Helminths (Worms in General) Most agents that mature in humans are helminths. Helminth diseases consist largely of those acquired by eating inadequately prepared seafood products usually infected with platyhelminths and nematodes, but occasionally with other agents. Some agents are freshwater parasites that, because of their complicated life cycles, have one or more stages that enter the coastal or marine environment. Presently, several of these are treated as freshwater cases because of the epidemiological aspects of the case. As more cases are studied, more human infections that involve strictly marine parasites will be demonstrated.

Many more cases of marine helminth infections probably exist than are reported or recognized by physicians. Surveys involving fecal examinations

exhibit adult worms or their eggs, but, for a variety of reasons, the samples are difficult to acquire and the corresponding infections are difficult to correctly diagnose or get reported. Most medical attention is not likely to culminate in confirmed reports of identified agents because physicians or technicians not familiar with proper techniques, means to correctly identify the agents, or people to consult are interested in treatments and not the specifics of a parasitic infection. Even when treated, the patient never knows what was causing the signs of disease because the agents are killed and digested or degenerate. On the other hand, in some countries where considerable raw or insufficiently prepared seafood is consumed, attention by researchers and the medical community results in an abundance of case reports. The Japanese and Koreans provide good examples of documented cases of infections by marine parasites.

Platyhelminths consist mostly of hermaphroditic soft-bodied worms. Those hermaphrodites of medical marine significance are cestodes (tapeworms) and trematodes (flukes), and most of those that mature in humans were consumed as larval stages. For many, the taxonomic identification is confused and the life cycle is partially or fully unknown. Moreover, infections are often undiagnosed because fecal examinations are not always conducted or are conducted by those without parasitological training. Moreover, the eggs of many species are similar in appearance.

Cestodes Cestodes infective to humans from marine hosts are members of the family Diphyllbothriidae in the Diphyllbothriidea, previously classified in the artificial assemblage Pseudophyllidea, now represented as comprising two unrelated clades, “Bothriocephalidea” and “Diphyllbothriidea” [6]. *Diphyllbothrium latum* is the best known member of the group, and it is considered a freshwater member with a cyclopoid-freshwater fish-bear/human/vertebrate cycle. Unpublished studies attempting to infect Mississippi marine fishes with the agent were unsuccessful; most copepod species developed a strong cellular response to the larva, and a mixture of “infected” copepods did not produce an infection in the Atlantic croaker. However, five human infections reported from eating the estuarine mullet *Liza haematocheila* in Korea were identified as *D. latum* on the basis of morphological characteristics [7].

Molecular tools should be able to confirm that identification as those tools have confirmed or corrected identifications of other species mentioned later.

Many human case reports have been identified as *Diphyllbothrium pacificum*, which typically matures in the pinnipeds. A recent taxonomic assessment of *D. pacificum* [8] involved numerous available specimens of variously misidentified species, including *D. latum* and *D. pacificum*. As it turned out, the classic publication [9] reporting human infections by *D. pacificum* in Peru was actually dealing with *Diphyllbothrium arctocephalinum*. Both *D. pacificum* and *D. arctocephalinum* occur along the South American Pacific coast in fur seals and secondarily in humans, with *D. pacificum* ranging north into Alaska and Japan and *D. arctocephalinum* restricted to the Southern Hemisphere, ranging south to South Australia and South Africa [8]. In one review, there are ten reported species of *Diphyllbothrium* from humans [10], including 6 from Japan [11], and in another there are 15 species [12]. Of the latter 15, not including *D. arctocephalinum*, there are eight nominal species in marine final hosts that also have been reported from humans: *Diphyllbothrium cameroni* (Hawaiian monk seal), *Diphyllbothrium cordatum* (Arctic seals, walruses), *Diphyllbothrium hians* (Arctic seals), *Diphyllbothrium orcini* (killer whale), *D. pacificum* (sea lions, eared seals), *Diphyllbothrium scoticum* (leopard seal, southern sea lion), and *Diphyllbothrium stemmacephalum* (harbor porpoise, bottlenose dolphin).

The species commonly infecting humans in Japan and previously referred to there as *D. latum* has been confirmed as *Diphyllbothrium nihonkaiense* ([8, 12], and others) (Fig. 1). The brown bear (*Ursus arctos*) is its natural definitive host, but it commonly infects humans, with some preserved worms measured as long as 8.8 m. Since infections are acquired from Pacific salmon (mainly cherry, pink, and chum salmon plus the Japanese huchen), either from the sea or just leaving the sea, it is considered a marine infection. Molecular studies have shown that human infections in Far East Russia up to the Kamchatka Peninsula reported as *Diphyllbothrium klebanovskii* are conspecific, identical to *D. nihonkaiense*. Molecular sequences also suggest the existence of two genotypes, biologically homogeneous populations not defined by locality [13]. Infections also involve consumers in Canada [14].



Waterborne Parasitic Diseases in Ocean. Figure 1
Author in Meguro Parasitological Museum, Tokyo, Japan, standing next to an 8.8-m-long specimen (> 10 m before fixation) of the tapeworm *Diphyllobothrium nihonkaiense* and the curators, the late Shunya Kamegai on right and Jun Araki on left

Moreover, salmon are exported to many countries, resulting in genetically confirmed infections in consumers in France and Switzerland [15] as well as New Zealand [16]. Another “freshwater” species, *Diphyllobothrium ursi*, also infects salmon and both brown and black bears as well as humans and dogs that feed on salmon or eat salmon-liver paste [17]. Cestodes and other parasites maturing in terrestrial definitive hosts that infect anadromous fishes such as Pacific salmonids or their freshwater fish or invertebrate hosts can be eaten by marine fishes, which in turn, serve as paratenic hosts. On the other hand, freshwater fishes can become paratenic hosts of helminths of marine origin. They can feed on infected marine fishes entering brackish river mouths or lower reaches of the rivers, or they can feed on anadromous fishes migrating upstream. A confirmed example is the plerocercoid of *Diphyllobothrium alascense* of dogs and rarely humans from the anadromous boreal smelt (*Osmerus mordax*) being transferred to the paratenic freshwater burbot (*Lota lota*) [18].

Members of *Diplogonoporus* have been reported from over 200 people in the twentieth century. Most cases have been identified as *D. balaenopterae*, or the names are probably synonyms of that species, a species

that infects consumers in Japan but also Korea [12, 19]. A case from Spain probably represents infection from fish imported from the Far East [20]. Genetic analyses of 18S rDNA, ITS1, and *cox1* nucleotide sequences of isolates were obtained from five Japanese cases [21]. The phylogenetic analysis of that material revealed little divergence and a close relationship to *Diphyllobothrium stemmacephalum*, a species maturing in the harbor porpoise and bottlenose dolphin rather than in whales as in *D. balaenopterae*. Data suggest that not only is *Diplogonoporus* paraphyletic, but those genetic data show that so are *Spirometra folium* and *Spirometra decipiens* [22].

Coastal Peruvians of ancient times had infections of *Diphyllobothrium* sp. as determined from coprolites [23], but infections of *D. pacificum* recently seem to be less prevalent in the northern Pacific Ocean, resulting from decreasing numbers of marine mammals and commercially important fishes like the pollack [8].

Not only have diphyllobothriid species been misidentified in the past, but human infections have been and will probably continue to be missed in many cases. They are most easily detected when a human expels lengthy portions of senile strobilae. Signs are variable but usually not severe [11, 24]. Most cases are probably not reported. The author did not experience any symptoms from an infection of four long specimens of *D. latum* other than mild discomfort several hours prior to periodic release of senile strobilae. The treatment of choice for the adult tapeworm infections in humans is a single dose of praziquantel (10–20 mg/kg) plus 30 g of magnesium sulfate as a purgative, if specimens are to be adequately studied.

Trematoda (Digenea, Flukes) Trematodes have a molluscan first intermediate host from which free-living cercariae are shed. The cercaria of most groups encysts or becomes encapsulated as a metacercaria in a second intermediate host, and this metacercaria matures as a hermaphroditic individual in a vertebrate final host when the second intermediate host is eaten. Because there are so many trematode taxonomic groups and species, there are numerous exceptions to the basic life cycle, and some of these involve species that can infect humans. For example, the well-known freshwater sheep liver fluke, *Fasciola hepatica*, encysts as small white spherules (metacercariae) on vegetation. Most people

that get infected do so by eating raw watercress on which the metacercariae have encysted. Schistosomes do not encyst but develop through the juvenile stage (schistosomula) in the final host where they ultimately mature in a blood vessel as separate male and female worms.

Depending on the taxon, the larval stage in the egg (miracidium) either hatches and penetrates its molluscan host or hatches after being eaten by a specific mollusk. The germ cells develop through a series of two or more asexual phases before shedding a continual production of hundreds to millions of infective cercariae. Some species have broad specificity in the molluscan, second intermediate host, or final host. Second intermediate hosts differ by family of trematode, but those discussed below include fishes, crustaceans, insects, mollusks, and none. Most of those marine/estuarine species reported from humans have a broad specificity of mammal or avian hosts. Most are rather small intestinal worms that necessitate a microscope to detect them unless hundreds are present, and each of these produces a continuous flow of a large number of small (25–30 μm long) eggs in the human feces for weeks or years. Most species are difficult to obtain in a healthy condition after medicinal treatment, are not transferred from the medical community to taxonomic parasitologists, are difficult to distinguish from one another, and produce minimal symptoms in light infections. Consequently, the number of human-infecting species and the prevalence and intensity of infections with marine trematodes are most certainly underestimated. Most of what is known has come from parasitologists within the medical community of Japan and Korea, where consumers eat an abundance of raw and “inadequately” prepared seafood products.

According to recent literature [25], an estimated 18 million people are infected with fish-borne intestinal trematodes. This was considered an underestimate, and it included some freshwater species. In the Republic of Korea, 17 species represent five families (Heterophyidae, Echinostomatidae, Plagiorchiidae, Neodiplostomidae, and Gymnophallidae) [26]. Of those, ten are heterophyids and, of those, seven are prevalent among residents who consume raw flesh from estuarine fishes in south and west coastal areas [25–27]. Those are *Heterophyes nocens*, *Heterophyopsis*

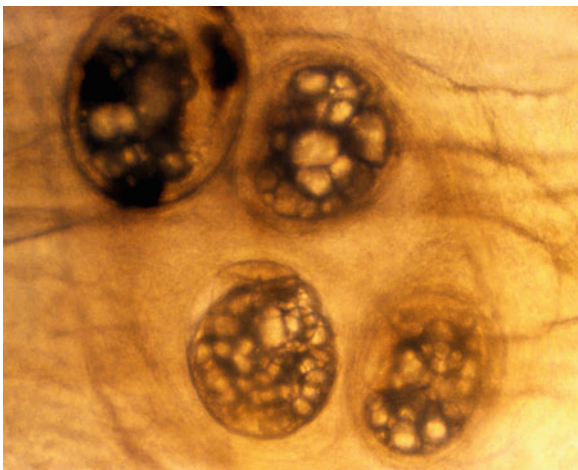
continua, *Pygidiopsis summa*, *Stellantchasmus falcatus*, *Stictodora fuscata*, *Stictodora lari*, and *Acanthotrema felis*. Documentation of all food-borne trematodes in humans includes 70 species in 14 families [25], but several are not from marine hosts and presently infected people are estimated to surpass 50 million.

Heterophyidae (Opisthorchioidea). As with zoonotic infections from eating fishes in Japan, Korea, China, and Thailand, such infections from Vietnam are now being recognized as a food safety risk in a country whose people have a strong tradition of eating raw fish [28]. The heterophyids use a snail first intermediate host, a fish second intermediate host, and either a bird or mammal definitive host. A prevalence of heterophyid infections with *Haplorchis pumilio* (100%), *Haplorchis taichui* (70%), *Haplorchis yokogawai* (6%), and *S. falcatus* (3%) was determined from expelled (25 mg/kg praziquantel, followed by a saturated solution of magnesium sulfate) adult worms in a survey of residents from coastal Nam Dinh Province south of Hanoi in 2005. The recent finding of those parasites in Vietnam suggested to the surveyors [28] that some may be introduced rather than endemic, increased in magnitude with intensification of aquaculture, or increased because of increased consumption and importation of raw fish. Difficulty presently exists in determining from which fishes the infections are acquired in people throughout Vietnam as well as the rest of the world where people eat raw fishes. Human-infecting species are beginning to be recognized in food fishes as those fishes are examined for parasites. For example, examination of wild and cultured grouper (*Epinephelus bleekeri*, duskytail grouper, and *Epinephelus coioides*, orange-spotted grouper, both marine species) have revealed *Heterophyopsis continua* and *Procerovum varium* (recognized as a freshwater parasite), and examination of mullet (*Mugil cephalus*) have revealed *Pygidiopsis summa* and *H. continua* from brackish water in Khanh Hoa Province in central coastal Vietnam [29]. The long tradition of eating raw fish has intensified with increasing affluence; a higher prevalence between men (69%) and women (23%) can be explained because the social gatherings where consumption of raw or pickled fish occurs have historically been male-oriented [28]. Concern exists

for the exportation of infected fresh fishes, some freshwater species, because aquaculture in rural Vietnam is increasing in economic importance [30].

Human infections have been known from Japan for decades even though most are freshwater species [25]. All three species of *Metagonimus* are freshwater species in regard to the snail host, but, in Korea, one, *Metagonimus yokogawai*, infects a marine fish, which consumers eat raw. Actually, an amphidromous fish, the sweetfish, *Plecoglossus altivelis*, spawns in rivers near the sea. Then, the fish fry feed on plankton at sea until 5–7 cm TL at which time they enter into the river mouths. Infections can be acquired from the fish in estuarine habitats. Sequence data readily separate the three species [31].

In the Western Hemisphere, there are several heterophyids that present a human health risk. For example, in the USA, species of *Phagicola* represent one of several potential infectious genera (not accepting *Phagicola* as a junior synonym of *Ascoctyle*). Knowledge about human infections from the Southeast US has not progressed much beyond anecdotal mention [1], but *Phagicola nana* (Fig. 2) from several fishes [32], *Phagicola longa* [33] from mullets (Fig. 3), and several other species (Fig. 4) probably occur in many consumers who eat inadequately prepared fish.

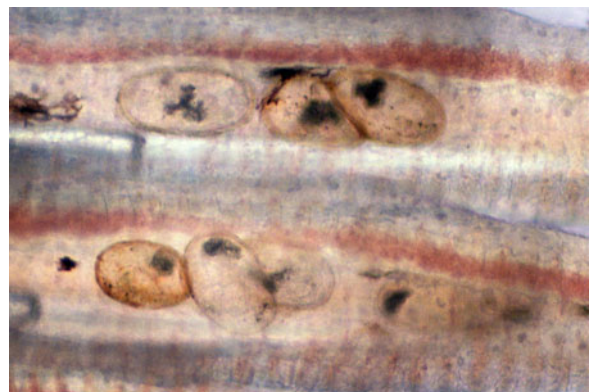


Waterborne Parasitic Diseases in Ocean. Figure 2
Living encysted metacercariae of the heterophyid *Phagicola nana* in fillet of largemouth bass, *Micropterus salmoides*, from Mississippi estuary. The cysts average 0.30 mm across

Heterophyes heterophyes, a heterophyid from the Middle East has been reported from fish-eating patients in Florida never having left the USA [34, 35]; these infections were most likely species of *Phagicola*. Actually, even capillariid eggs get misidentified as those of *H. heterophyes* in Egypt, where the trematode is common [36]. Unlike many heterophyids in the USA that exhibit a high “site specificity” in the second intermediate fish hosts, meaning found in one or few specific organs or tissues, *P. nana* occurs in many



Waterborne Parasitic Diseases in Ocean. Figure 3
Mounted and stained adult specimen of the heterophyid *Phagicola longa* from an experimental infection originating from a metacercaria in *Mugil cephalus* from Mississippi

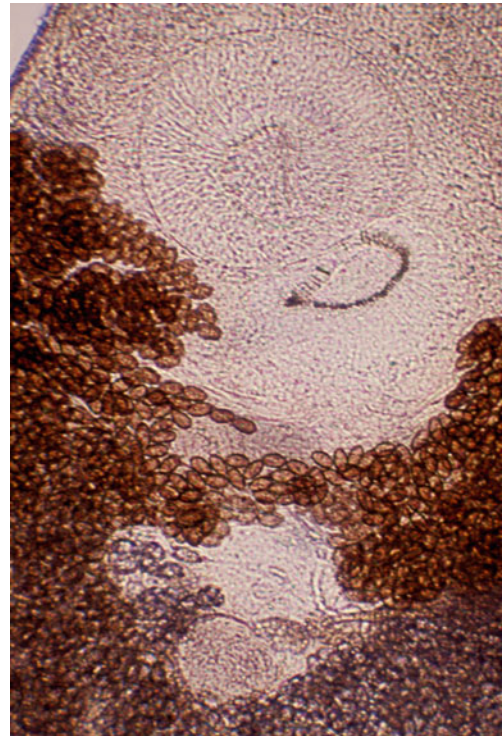


Waterborne Parasitic Diseases in Ocean. Figure 4
Living encysted metacercariae of the heterophyid *Phagicola diminuta* in gills of Gulf killifish, *Fundulus grandis*, a bite-sized fish from Mississippi estuary

tissues such as the muscular fillets in addition to the viscera, gills, and other sites [32]. Moreover, *P. nana* infects a few different fishes such as the largemouth bass (*Micropterus salmoides*) and several sunfishes (Centrarchidae), popular recreational fishes in estuaries of the Southern US that are consumed in a variety of cuisines. Most heterophyid species infect specific organs in specific small fishes that are eaten raw in their entirety by specific groups of people, if eaten at all. As with several different parasites, some heterophyids that constitute public health risks constitute species complexes that are difficult to distinguish by morphological features. For example, what is commonly reported as *Centrocestus formosanus* is a mixture of morphologically similar species. Members of this complex are common in the Far and Middle East, but some apparently have been introduced into Hawaii, USA mainland, and Mexico through introduced infected snails (*Melanoides tuberculatus*) in the aquarium trade on vegetation and with fishes [37]. Members of this basically freshwater heterophyid complex infect a large number of fish hosts, including some estuarine species.

The reason for stressing heterophyids is because, even though most species measure 1 mm long or less, they often occur in large numbers, influence public health, and can result in severe pathological changes in humans. A series of studies in the 1930s based on human infections in the Philippines was summarized in 1940 [38]. Five heterophyids were identified from 33 of 297 autopsy cases, not all randomly chosen: *H. yokogawai* (16 cases, 13 alone), *H. taichui* [5], *Haplorchis calderoni* [2], *Haplorchis vanissima* [1], and *Diorchitrema pseudocirrata* [6]. These occurred either singly or with other species. The other species consisted of the microphallid *Microphallus brevicaecum*, which was originally identified as a heterophyid (*Heterophyes brevicaeca* but later recognized as *Spelotrema brevicaeca*) (in 11 cases), an echinostomid identified as *Euparyphium ilocanum* [2], and an unidentified plagiorchiid (*Plagiorchis* sp.) [1], plus the common heterophyids already listed. All the trematodes occur normally in the intestine, where lesions were not found to be severe. More important, eggs of the heterophyids *H. yokogawai* and *H. taichui* and the microphallid *M. brevicaecum* were also associated with acute and chronic lesions in the heart and other visceral organs such as liver, spleen, kidneys, brain, spinal cord, and lungs. *Diorchitrema*

pseudocirrata also occurred in some of those lesions in mixed infections but never by itself. Worm eggs, especially from degenerated adults, enter into the blood and lymphatic streams and are carried to different visceral tissues where they provoke the cellular reactions. Adhesions and hemorrhaging were common complications. Of 34 positive cases (1 did not exhibit an intestinal infection), 15 (44%) had visceral complications (1 in brain with fatal hemorrhaging, 1 in spinal cord, and 13 in heart). Of 89 enlarged hearts in these charity patients in the Philippines, the 13 (15%) with myocardial or valvular lesions associated with worm eggs were assumed to have caused the heart dysfunction, a value attributed to heterophyid-caused fatal heart disease [39]. Adult worms (Fig. 5) identified as *Heterophyes heterophyes* also encapsulate in the human brain [40].



Waterborne Parasitic Diseases in Ocean. Figure 5 Close-up of living 0.8-mm-long adult specimen of the trematode *Heterophyes* sp. from Bardawil Lagoon, Northern Sinai, Egypt, showing diagnostic sclerites in the modified genital sinus, often misstated as a "gonotyl," below the ventral sucker

An enormous number of infective metacercariae can occur in some habitats. To appreciate the life cycle, each adult out of the thousands that can be present in the intestine of a human or in that of a domestic dog or cat produces thousands of eggs, with several hundred laid daily for several weeks or months. One dog was reported to host 13,000 specimens of *H. heterophyes*. Each egg produces a miracidia that can infect a snail. In the lakes along the Nile Delta and in the Bardawil Lagoon, over 90% of the host snail population can be infected (Figs. 6, 7). Just short of 500 cercariae can be released from each of the

abundant snails daily for a year, but with decreasing cercarial production over time. Each cercaria survives 1–2 days, and mullets are common hosts. Eleven emaciated specimens of *Liza ramada* contained 1,730–6,000 metacercariae per gram for an estimated 582,000 metacercarial cysts in one 255-mm-long fish. A specimen of *Mugil cephalus* from the same location had 1,136 metacercariae per gram [41]. Fishermen, seafood dealers, children of those people, and those who taste fish being prepared commonly become infected because the rules established for protection of the consumers buying the products are not always followed by the handlers. Moreover, fishermen do not always completely cook their fish.



Waterborne Parasitic Diseases in Ocean. Figure 6
Street vendors selling fresh fish in El Arish, Northern Sinai, Egypt



Waterborne Parasitic Diseases in Ocean. Figure 7
Close-up of activities in Fig. 6, showing vendor displaying salted mullet. A regulation at the time was to keep fish salted for 8 days before selling them to the public. Eight days was necessary for the heavy infections of heterophyid metacercariae to lose their infectivity

Microphallidae (*Microphalloidea*). Members of small intestinal trematode complexes other than Heterophyidae also have been misidentified. Microphallidae constitutes one of those families. One or more microphallid members also have a relatively broad specificity of the final host and can also cause severe potentially fatal pathologic lesions [38, 42]. As mentioned above, *Microphallus brevicatum* from shrimp has been determined to be the cause of human mortalities in the Philippines. Microphallids are superficially similar to heterophyids and also infect birds and mammals and occasionally fishes, reptiles, and amphibians, but the microphallids have entirely different cercariae, differ phylogenetically, and use crustaceans such as crabs and palaemonid and penaeid shrimps as second intermediate hosts rather than fishes.

Troglorematidae (*Gorgoderioidea*). *Nanophyetus salmincola* is another small, about 1-mm-long, trematode that infects the intestine of numerous mammals, including humans. In fact it is probably unknowingly the most common helminth in people of the North American Pacific Northwest. It produces gastrointestinal distress, fatigue, loss of weight, and peripheral eosinophilia when numerous specimens occur in humans, and can be diagnosed easily from its relatively large, 0.090–0.100-mm-long, diagnostic eggs that appear in stools about a week after eating infected fish. No symptom occurs in most light infections. Basically a freshwater parasite, cercariae form metacercariae in several salmon and steelhead trout (*Oncorhynchus mykiss*, also known as sea-run

rainbow trout) as well as freshwater fish and amphibians that do not make oceanic migrations. Infections in the Pacific Northwest are attributed to eating raw, incompletely cooked, cold-smoked salmon and steelhead trout and their eggs [43]. A second species, *Nanophyetus schikhobalowi*, needs to be studied molecularly to determine if it is conspecific with *N. salmincola*. It had been demonstrated to infect as many as 98% of the people in some eastern Siberian villages.

The trematode in the Pacific Northwest serves as host for the rickettsial agent, *Neorickettsia helminthoeca*, which in turn causes “salmon poisoning disease” of canines. Dogs that are fed infected salmon may die, but humans and other non-canines are not known to be affected. Nevertheless, the loss of dogs in the Arctic can be a serious problem.

Gymnophallidae (*Gymnophalloidea*). *Gymnophallids* mature in shore birds, gulls, and diving ducks. However, members of *Gymnophalloides* also infect mammals. Human infections with *Gymnophalloides seoi* are known only from Korea, but most are tiny, <1 mm, and it and other species all use a variety of mollusks as first and second intermediate hosts and probably infect people throughout the world where mollusks are consumed raw. *Gymnophalloides seoi* is acquired primarily from eating the raw oyster, *Crassostrea gigas*, known as the Japanese oyster, Miyagi oyster, giant oyster, immigrant oyster, and giant Pacific oyster, and infects the intestine and possibly pancreatic duct of human consumers [25, 27]. Nevertheless, like *G. seoi*, which infects the Palearctic oystercatcher, *Haematopus ostralegus*, most members of this genus can be experimentally administered into a variety of birds and mammals, suggesting the human health risk. Studies with mice demonstrate that *G. seoi* produces focal responses for 2–3 weeks, at which time the mucosal integrity is restored. If provided with immunosuppressive treatment, the mice underwent minimal goblet cell hyperplasia, extended retention of worms, and invasion of the submucosa. Even though genetics of hosts also play a role in this sequence, immunosuppressed humans are at risk for transference of eggs and worms to remote organs as with heterophyid infections. With a relatively low production of eggs relative to heterophyids, a total of

100 specimens would be necessary to produce 8,400 eggs per day, considerably fewer than most heterophyids [25].

Echinostomatidae (*Echinostomatoidea*). Members of some other trematode families also infect the small or large intestine but grow slowly and reach larger size (roughly 3–7 mm) compared with those mentioned above (0.5–2.0 mm); a few marine members are known to infect humans. A total of 21 members of *Echinostomatidae* are listed as causing human infections, some resulting in gastroenteritis [25]. Of those, all were from freshwater mollusks except two. One, *Acanthoparyphium tyosenense*, can be acquired in Korea from eating at least the raw bivalves *Macrta veneriformis* and *Solen grandis* and the gastropod *Neverita bicolor* available at a coastal village seafood market [44]. The gastropod first intermediate hosts *Lunatia fortunei* and *Glassaulax didyma* shed cercariae that infect the gills of the popularly eaten bivalves. Ducks serve as the natural definitive host, but experimental infections in chick and seagull have confirmed the adult infection [25]. Related species, and there are several, such as what was identified as *Acanthoparyphium spinulosum* in the oyster *Crassostrea virginica* from Texas, USA, are probably also commonly infective to humans, but such infections are limited only by the infrequency with which the molluscan second intermediate hosts are eaten raw [45].

Plagiorchidae (*Plagiorchioidea*). Human infections with *Plagiorchis* spp. are uncommon, but 12 cases have been reported. *Plagiorchis vespertilionis* occurred in a man from coastal Republic of Korea who often ate raw mullet and gobies and was not known to eat caddisfly larvae, mayfly larvae, or dragonfly nymphs. The insects were suggested as possible intermediate hosts and are known hosts for many other species of *Plagiorchis*. The patient also contained *Heterophyes nocens*, acquired from fish [46]. Nevertheless, fish were suggested intermediate hosts, and perhaps the infection was transmitted by eating raw estuarine fish that had recently fed on infected dragonflies or some infected insect larva, joining the freshwater parasite with the coastal habitat. Members of *Plagiorchis* and related genera mature in birds and mammals plus accidentally in amphibians and reptiles. They are

relatively large (3–5 mm) and usually occur in low numbers in the final hosts.

Didymozoidae (Hemiuroidea). Humans do not get infected by didymozoids! These trematodes of fishes are unusual in that they occur usually in pairs of either hermaphroditic individuals or as separate sexes like mammalian schistosomes and occur encysted in the flesh, encysted in other tissues, or even free in body spaces other than the lumen of the alimentary tract like most trematodes. The group is mentioned because people eat infected fish products and diagnosticians find the small reniform didymozoid eggs (roughly 0.015–0.030 mm long but in some species up to 0.040 mm) of the worms in human stool samples, leading those diagnosticians to assume either an active trematode or protozoan infection. Such misidentifications have been reported from people eating marine fishes in Japan, Taiwan, and the Philippines, but cases could occur throughout the world where infected fish are abundant, especially in pelagic tropical and subtropical fishes [47].

Nematoda (Roundworms) Nematodes, or roundworms, comprise both free-living and parasitic groups. Parasitic groups that have marine members exhibiting a public health risk also fit into different groups: those that mature in humans, those that have larvae that can infect humans eating inadequately prepared seafood, those that produce excretions or secretions that result in an immunological response in humans, and those that can be acquired by being in the water with free-living stages or handling infected products. When consumers think of parasites in seafood, they typically think of anisakiasis involving grossly apparent nematode juveniles. Of the many nematode groups, each may have a different strategy in completing a life cycle. An example cycle will be mentioned for each major group of worms.

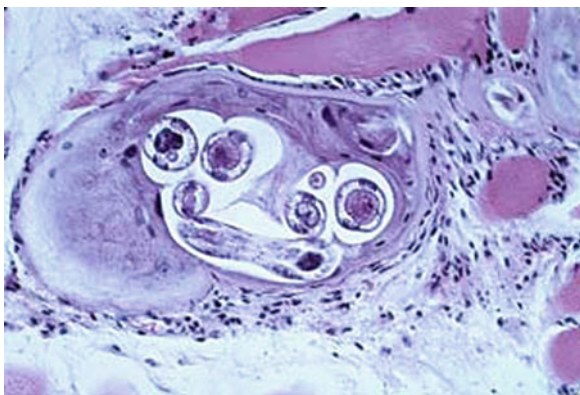
Trichinellidae (Dorylaimia, Trichinellida). Trichinellosis has received considerable attention over the last 150 years of control efforts, but in the past 20–30 years, there has been a dramatic reemergence as an emerging zoonosis [48]. The reasons are severalfold involving changing agricultural, marketing, environmental, and economic practices and conditions, but basically prior attentions involved *Trichinella spiralis*

in domestic pork. In the 1940s and 1950s, a large number of marine and terrestrial mammals in Alaska were examined for what was then known as *T. spiralis*, and most had infections [49]. A total of 320 native individuals from three Alaskan areas given intradermal tests showed that 28% from Wainwright, Alaska, tested positive. There are now at least ten recognized species or genotypes, and the involvement of the sylvatic cycle of some of those is responsible for new human cases. For purposes of this review, *Trichinella nativa* and to a lesser degree *Trichinella britovi* play a role. *Trichinella nativa* infects the carnivores that inhabit the marine environs: polar bear (*Ursus maritimus*), walrus (*Odobenus rosmarus*), Arctic fox (*Alopex lagopus*), and wolf (*Canis lupus*), but it is even known from China in domestic pigs. This species differs from the others in that it can survive the freezing process, the process that keeps domestic pork safe to eat. Uncertainty exists in how long the frozen juveniles remain infective, but, when juveniles from walrus meat were stored at -20°C for up to 20 months, some remained infective to experimental guinea pigs [50]. There was a gradual degradation process that rendered some noninfective. The freezing resistance appears to be related to the parasite-nurse cell-host complex, and the nurse cells yield a cryoprotectant when occurring in carnivore muscle tissue but not when in rodents [51]. Outbreaks are common in the Canadian Arctic where uncooked walrus meat is eaten raw, frozen, and aged by Native Americans as “igunaq” [52], but some that get infected also eat the products cooked [53]. This product is produced by placing meat and fat tissues into skin bags, which are sewn shut and aged about a meter under a cache of rocks or gravel along a beach. When meat from experimentally infected seal was prepared as igunaq and allowed to ferment, as air-dried meat (nikku), and as both raw and partially cooked sausage (core temperature less than 50°C), it remained infective for at least 5 months under laboratory conditions. The food product was fed to cats and the extracted juveniles were orally inoculated into mice [54]. Whale is usually eaten as muktuk, a preparation of fermented skin and blubber containing no muscle [53].

Trichinella britovi also occurs in temperate areas of the Palearctic region where people obtain infections primarily from eating wildlife. Apparently, it cannot resist freezing for a long period; the juveniles of

T. spiralis, which have been reported from a variety of marine mammals, die after 4 days at -10°C . The status of infections from marine mammals requires additional research. The cycles of all the species are similar to each other but different from other nematodes in this chapter. The definitive host, including human, eats uncooked infected meat from another definitive host, and the juveniles are released when the cyst is digested in the stomach. The juveniles invade the mucosa of the upper intestine and develop into mature adults within 2 days. After 4 days, the mated female starts depositing motile juveniles, a process that continues for 4–16 weeks and produces up to 1,500 juveniles in the nonimmune host. These juveniles pass within the lymphatic system or mesenteric venules to muscles, usually striated ones, where it encysts. Striated muscle tissue stimulates development of nurse cells and capsule formation within 2–3 weeks (Fig. 8). The coiled juveniles can survive in human muscle for years, even though calcification may occur within 6–9 months. Infections can be fatal with as few as five juveniles per gram of muscle, but infections can be much higher, especially if the initial dose in an immunologically naïve consumer contains hundreds of cysts in striated muscle. A low initial dose establishes a protective immunity to challenge doses.

Paracapillaria philippinensis (until recently [55], a trichinellid that was and still is referred to by some as *Capillaria philippinensis*) seems to be spreading geographically. This small intestinal nematode produces



Waterborne Parasitic Diseases in Ocean. Figure 8
Section of coiled specimen of juvenile *Trichinella spiralis*
encapsulated in striated muscle tissue

mortality in humans. More than 2,000 human cases have been reported from the Philippines and Thailand, with several of those from more recent cases in Korea, Japan, Taiwan, India, Iran, United Arab Emirates, Egypt, Italy, Spain, and the United Kingdom [56], but with some of those acquired from eating freshwater fishes or imported fishes. The number of human cases is being reduced in Thailand, but infections probably contribute significantly to mortality there [57]. The life cycle probably includes a bird definitive host as shown by experimental infections in a fish-eating rail and heron, even though many mammals can host transient infections, and humans and the Mongolian gerbil can die from infections in less than 2 months [58]. Eggs passed by the definitive host hatch in a variety of freshwater and estuarine fish species. At least when the gerbil or human eat the infected fish, the juveniles develop into males and females that release a first-stage juvenile from a thin-shelled egg, which in turn can autoinfect the host and produce adults, with a female that produces the typical capillariid thick-walled eggs that are released in the feces. Even though most second generation females are oviparous, a few females still produce juveniles that can progress to autoinfection and hyperinfection, a condition in which the produced juveniles penetrate the intestinal mucosa and proceed with migration. At autopsy, 1 l of bowel fluid contained an estimated 200,000 specimens in all stages [58].

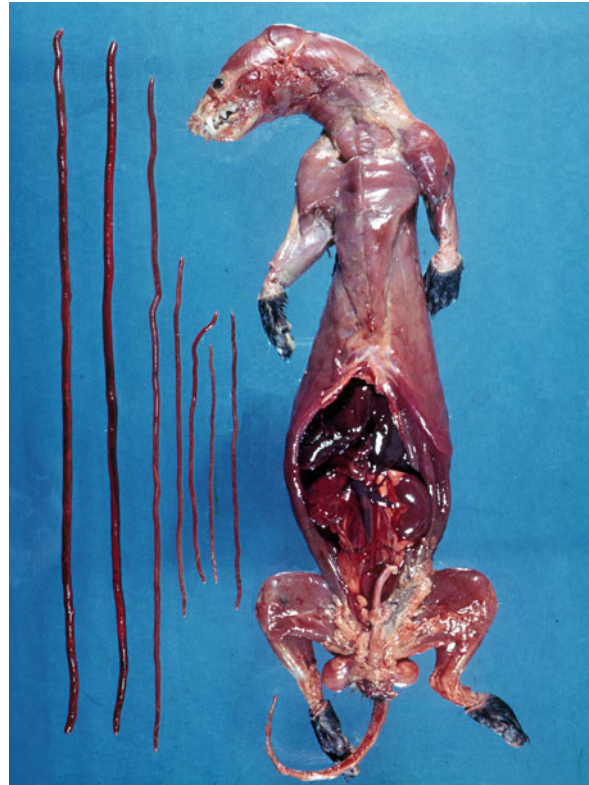
Signs of infection in humans are varied: they can include abdominal pain, diarrhea, vomiting, heart irregularities, edema, and weight loss, with the blood exhibiting low protein, calcium, potassium, and sodium, and with high immunoglobulin E (IgE) with diminished levels of IgG, IgM, and IgA for several months after treatment with albendazole. Albendazole seems more effective than thiabendazole and mebendazole because there is no relapse, presumably because only albendazole affects the larval stages [58].

Epidemiological studies are lacking in most reports, but where infections are common, elderly aboriginal groups with infections in Taiwan prefer Chou-Bao, which includes soaking raw, small entire fish in fermented millet [56], and infected Filipinos consume Kinilaw, which consists primarily of raw pilchards, or other small fishes from coastal lagoons, marinated in vinegar, garlic, onions, ginger, tomato, and various

peppers. The small fishes throughout SE Asia also are eaten whole as raw products, and some native groups (Ilocanos of Northern Luzon, Philippines) use animal organs and intestinal juices to season rice and other foodstuffs [58]. Most human cases have been traced to consumers that eat small, raw entire fish, but exact causes are not known. When diagnosis is delayed, the disease can be severe, with up to 30% fatality in some localities in the Philippines.

Diectophymatidae (Dorylaimia, Diectophymatida). The giant kidney worm, *Diectophyme renale*, rarely infects humans, but the potential because of aquaculture and new cuisines is increasing. Even though usually considered a freshwater parasite, it also is common in the estuarine environment [1]. Its life cycle involves an oligochaete and a fish, amphibian, or reptilian paratenic host. Its counterpart *Eustrongylides* will be discussed later because it does not mature in humans. *Diectophyme renale* matures in and completely devours the kidneys of its definitive host. The site is usually restricted to the right kidney, probably because of migratory behavior of the acquired juvenile. If it does not encounter the right kidney, it can mature in the peritoneal cavity, where it also causes harm (Fig. 9); it can also become encapsulated and form a subcutaneous nodule [59]. Human infections are usually diagnosed by the worm eggs in urine, followed up by radiological exams to detect enlarged or calcified kidneys. The worm has a cosmopolitan occurrence, but most human infections are from Iran and from freshwater habitats. In fact, human coprolites dated about 3,375 YBC from an archeological site in Switzerland suggest that human infections were more prevalent in the Neolithic than now [60]. In the USA, sylvatic infections are common in the mink (*Mustela vison*) in brackish marshes of southern Louisiana [1].

Acanthocephalans (Spiny-Headed Worms) Acanthocephalans from marine products are not known to mature in humans. However, a few from terrestrial arthropods such as cockroaches do produce eggs in humans. In fact, eggs were found in human coprolites from a cave in Utah dated about 1869 BC $\pm$ 160 years and 20 AD $\pm$ 240 years [61]. On the other hand, juveniles can attach in the lumen of the intestine without maturing, and, because they are in the lumen and can be voided in the feces, they are treated here.



Waterborne Parasitic Diseases in Ocean. Figure 9 Specimens of the giant kidney worm, *Diectophyme renale*, removed from a mink's abdominal cavity in Louisiana. This species usually devours all the parenchyma in the right kidney of the mink and other fish-eating carnivores, but these specimens ended up in the cavity and matured without killing the host

This seems to be the case for *Corynosoma strumosum* in a person from Alaska [62] and *Bolbosoma* sp. from a Japanese fisherman [63]. The first case was noticed when the infected male was treated with Atabrine. The worm matures in seals but is known as a juvenile in dogs, otters, birds, and other fish-eating birds. When fresh marine fish were fed to mink, some mink died [64]. Even though not routinely detected in people, *C. strumosum*, common as adults in pinnipeds and as juveniles from birds and other mammals in the Holarctic, was considered probably to be common in humans but for a short duration [18, 62]. The second case was detected because the worm had penetrated the intestine and was surgically removed because the disease was initially diagnosed as appendicitis.

Acanthocephalans are like nematodes because they both have separate males and females. However, all species are parasitic, and molecular data place them with rotifers. The life cycle includes an arthropod or crustacean and a vertebrate. In the case of the two palaeacanthocephalans mentioned above, the crustacean is unknown, but it could be small or large. The life history is typically completed by seal, whale, or sea otter by feeding on a paratenic fish host containing an encysted or encapsulated larva or juvenile. Most acanthocephalans have a rather specific crustacean and final host and a larger but still rather specific range of paratenic hosts. Even though the definitive host is quite specific, the worm may attach in the intestine of non-closely-related hosts. For example, two palaeacanthocephalan species that mature in a few teleost fishes can be acquired by local dasyatid stingrays as juveniles from amphipods and attach to the spiral valve without developing further or without migrating to the viscera or muscles as paratenic parasites [65]. Stingrays are not typical final hosts for any acanthocephalan.

Information on additional juvenile acanthocephalans that penetrate through the alimentary tract and cause a host response, sometimes severe, appears later in the next section of this chapter.

Parasites with Larvae That Can Infect Humans Eating Inadequately Prepared Seafood

The best-studied parasitic infections in humans eating raw marine seafood are caused by members of the ascaridoid genus *Anisakis* and first recognized in the 1960s in the Netherlands, Japan, and the USA, even though Native Americans in Alaska passed specimens of both *Anisakis* sp. and *Pseudoterranova decipiens* sensu lato in feces [66]. There is now confusion in the names given the human disease because ascaridoids other than species of *Anisakis* also induce similar infections. These agents plus infections and potential infections by non-ascaridoid members of a variety of classes, based on the ability of such species to cause disease in nonhuman mammals, are also discussed in this chapter. In 1988, a group interested in standardizing nomenclature of ascaridoid diseases recommended calling disease involving a member of *Anisakis* as “anisakiasis” and disease caused by any member of the family Anisakidae as “anisakidosis,” even though

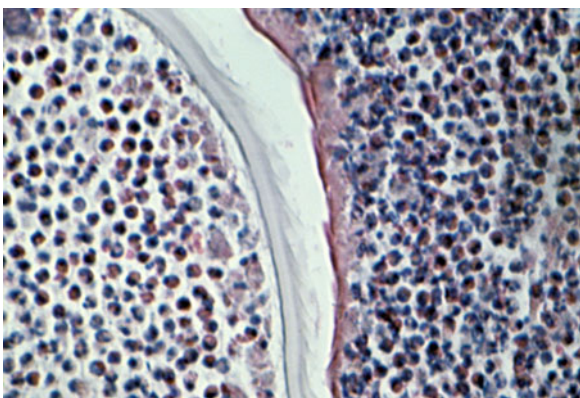
some authors use a name referring to the responsible generic agent or the host response, such as “eosinophilic granuloma.” DNA-based techniques have allowed accurate diagnosis, species differentiation, and phylogenetic relationships. Diseases caused by members of other higher taxa also create some confusion with terminology of the names.

Anisakidae (Spirurina, Ascaridomorpha). Species of *Anisakis* are the most widespread cause of notable and severe human disease. Until recently, human infections were divided into two types of juveniles, *Anisakis* type I and type II [67] based on the presence of a caudal spine and the length of the ventriculus, a glandular structure separating the muscular esophagus and the intestine. A group of 21 species based on adults were divided into three primary species [68], including *Anisakis simplex* with ten junior synonyms, *Anisakis typica* with one synonym, and *Anisakis physeteris* with three synonyms. With improved molecular tools, some of those junior synonyms have been accepted, and additional species have been recognized. *Anisakis* type I juveniles comprise the *Anisakis simplex* complex: *Anisakis simplex* sensu stricto, *Anisakis pegreffii*, *Anisakis simplex* C [69], *Anisakis ziphidarum*, and *A. typica*. *Anisakis* type II juveniles comprise *A. physeteris*, *Anisakis brevispiculata*, and *Anisakis paggiae*. All of these can be differentiated and are considered human pathogens. The differences in pathogenesis and location in the human probably reflects the different species, at least in part. Recent polymerase chain reaction amplification (PCR) methods show that *A. pegreffii* comprises the source of most infections in consumers in Italy but that *A. simplex* sensu stricto is the most common species in people in Japan (Figs. 10 and 11) even though fish in the Sea of Japan are known to be infected with *A. pegreffii* [70]. Also, *A. pegreffii* seems to cause gastric distress in humans and *A. simplex* seems to mostly affect the intestine. Exceptions involve the abdominal cavity, mesentery, and oral cavity [71]. Juveniles typically measure 1–3 cm in length.

The different species can be differentiated by cetacean host and geographical locality. *Anisakis schupakovi* infects the Caspian seal, *Pusa caspica*, and belongs in the genus as do a couple other species from delphiniids in South America rivers and other isolated habitats. Tabularized reviews list which species of *Anisakis* have



Waterborne Parasitic Diseases in Ocean. Figure 10 Section of *Anisakis simplex* in ileum of human patient in Japan. Note the characteristic eosinophilic response, referred to as “eosinophilic granuloma,” surrounding the diagnostic section of the worm



Waterborne Parasitic Diseases in Ocean. Figure 11 Section through degenerating specimen of *Anisakis* sp. in stomach of experimentally infected miniature pig showing juvenile with infiltrated eosinophils and surrounding infiltrate consisting almost exclusively of eosinophils

been confirmed molecularly from different cetaceans, cephalopods, and fishes [70]. The life cycle of *Anisakis* spp. includes a cetacean final host, from which worm eggs (Fig. 12) are deposited in feces and embryonate in seawater; the active, hatched third-stage juvenile (the same as for *Pseudoterranova decipiens*), also referred to as a larva, is eaten by a crustacean [72]. The juvenile hatching from most nematode groups is a second-stage juvenile. Euphausiids are known to host the infective third-stage juvenile. A variety of paratenic hosts, including specific fishes, squids, and crustaceans can acquire and maintain the infective juveniles by feeding on the euphausiid or some undetermined crustacean hosts. The appropriate cetacean feeds on any appropriate animal with an infective juvenile. In some cases, there can be a large cluster of these worms, involving third-, fourth-, and fifth-stage juveniles and adults attached within a stomach ulcer. As with several ascaridoids in a variety of genera and host groups, individuals not embedded in the ulcerated tissue will leave the ulcer and forage among the digesting prey in the stomach only to return at a later time.

Cultured fish usually do not exhibit ascaridoid infections present in their wild counterparts because they are fed a diet that does not contain infective



Waterborne Parasitic Diseases in Ocean. Figure 12 Egg of *Hysterothylacium reliquens* showing infective third-stage juvenile representative of ascaridoids infective to humans. In this species, the juvenile hatches and the free juvenile is eaten by an appropriate copepod, which in turn is eaten by an appropriate fish. That fish or the copepod is eaten by a paratenic host or by the definitive fish host

juvenile worms. Such is not always the case. Infected prey can enter a sea pen or be fed to the cultured fish in fish farms or cages. When juvenile specimens of the greater amberjack, *Seriola dumerilli*, and three-line grunt, *Parapristipoma trilineatum*, were imported into Japan from China for grow-out in fish farms, they were determined to be heavily infected with *Anisakis pegreffii* [73]. The Japanese Ministry suggested the fish were fed wild infected fish before being exported. In any event, it required that the products, originally grown for sashimi (raw fish), be frozen prior to taking them to market, considerably reducing their value.

Humans become infected with juveniles of *Anisakis* when they eat uncooked or inadequately prepared seafood containing an infective third- or fourth-stage juvenile. Symptoms and signs of infection probably vary by species of *Anisakis*, condition of worm, and condition and genetics of patient. Some patient complaints include sudden epigastric pain, nausea, diarrhea, and urticaria occurring within an hour but usually at about hour 6. Because infections are difficult to diagnose, the acute infection caused by a migrating or encapsulated juvenile develops into a chronic case with intermittent abdominal pain lasting from weeks to several years. There may be a slight fever and moderate leukocytosis up to 15,000 leukocytes per cubic milliliter. Peripheral eosinophilia can vary from 4% to 41% and will be discussed later in the chapter. The condition presenting other vague symptoms cannot be diagnosed with various immunologic assays because the sera cross-reacts with related nematodes or even with human sera [74], but more specific serological tests are being perfected and various molecular tools are being developed [75, 76]. Most prior cases have been identified morphologically after misdiagnoses of appendicitis, ileitis, cholecystitis, Crohn's disease, and tuberculosis peritonitis, and a variety of neoplasms resulted in resection of the corresponding lesions. In most cases, the worms penetrate or embed in the ileum or duodenum but occasionally penetrate the stomach as for *A. pegreffii*. Anisakiasis occurs in the Netherlands and other European countries, Japan, the USA, and Chile, and cases are remaining high. When the estimated prevalence of *A. simplex* (*A. pegreffii*?) -specific IgE assay in the population of Madrid, Spain, was tested by enzyme-linked immunosorbent assay (ELISA) as well as for cross-reactivity, 12% of

the healthy population tested positive compared with 22% in southern Spain determined with a somewhat different technique [76] and without any indication of cross-reactivity. Those testing positive in Madrid were more likely to admit to being habitual consumers of fresh (Fig. 13) or undercooked fish as opposed to eating frozen and cooked products like reported for the negative population.

As with "*Anisakis simplex*," "*Pseudoterranova decipiens*" also was recognized as a single species. It infected fishes and humans and could be identified by a morphological diagnostic juvenile and a variable adult, but the adult matured in pinnipeds rather than cetaceans. Recently, however, this agent of human infections that was supposed to occur worldwide has been shown to include several species, most in specific



Waterborne Parasitic Diseases in Ocean. Figure 13 Italian fisherman aboard fishing boat with catch including sardines infected with *Anisakis pegreffii*. Fresh fish from boats in the Port of Civitavecchia, Italy, are boxed, iced, and sold to restaurants, seafood dealers, and others

pinnipeds and consequently in specific geographic regions. Molecular and biochemical tools, backed with morphological features, have shown this species to represent a complex of at least six species: *Pseudoterranova decipiens* sensu stricto, *Pseudoterranova krabbei*, *Pseudoterranova bulbosa*, *Pseudoterranova azarasi*, *Pseudoterranova cattani*, and *Pseudoterranova decipiens* E. There are also exceptions to the strict host specificity within this genus: *Pseudoterranova ceticola* and *Pseudoterranova kogiae* both infect the cetaceans *Kogia simus* and *Kogia breviceps* (dwarf and pigmy sperm whales, respectively) [70].

When consumers eat infective juveniles of *Pseudoterranova* spp., the condition is usually not serious. Specimens are typically longer than those of *Anisakis* spp., roughly 4 cm versus 2 cm but dependent on age and development of worm, species, and temperature, and they usually have a more reddish or darker coloration. Whether they occur in the flesh or viscera of the fish depends on the species of worm and species of fish for both *Anisakis* spp. and *Pseudoterranova* spp. Specimens are often noninvasive and void in the feces or vomited; they tickle the throats of those eating inadequately prepared seafood when they migrate from the stomach back up the esophagus into the oropharynx, or they are coughed up into the mouth. *Pseudoterranova* spp. that cause disease usually involve the mucosal or submucosal tissues of the stomach. The common name of *P. decipiens* from the harbor seal had been “codworm” because the species was so abundant in cod, the second intermediate/paratenic host, especially during strict protection of harbor seal pups, which contained heavy infections. Actually, the “sealworm” consists of *P. decipiens* sensu stricto and *P. krabbei*, which can occur in mixed infections in *Phoca vitulina* and *Halichoerus grypus*, and both infect the cod, *Gadus morhua* [70]. Infections of large, 2–4-mm-long, unidentified fourth-stage juveniles of species of *Pseudoterranova* have become more prevalent and better recognized in Chile during recent years. Of 15 patients, all but one appeared to lose the infection through the mouth after irritation; one was voided through the anus with intense diarrhea. The patients all enjoyed seviche or fried fish [77]. Differing clinical signs between human infections in the USA and Japan may be different because the cause is from separate species. These may be *P. bulbosa* and *P. azarasi* in

Japan and *P. decipiens* sensu stricto and *P. bulbosa* in the USA (Fig. 14) [70].

The life history of *Pseudoterranova decipiens* sensu lato is more studied than that of known species of *Anisakis* sp. [78]. The life cycle differs slightly from that of known species of *Anisakis* sp. The first intermediate host is a small crustacean such as several mycidacean species that spend part of their life in the benthos [79]. The hatched third-stage juveniles are not active. The agent in the crustacean appears to require a fish or large invertebrate paratenic host to complete the history in the pinniped.

Members of *Contracaecum* have been reported from both pinnipeds and fish-eating birds. Even though morphologically they appear similar, those from the pinnipeds and birds are molecularly separated, perhaps with those species from seals representing *Phocascaris*, which is paraphyletic within *Contracaecum* [70]. Like for *Pseudoterranova decipiens*, *Contracaecum osculatum* from phocid seals consists of a complex of at least five genetically recognized, but not all named, species. There are 11 mammalian species that have been sequenced. Because they are from temperate and polar areas, there is less known about human infections, but they are discussed based on known genetic information [70]. Two human cases of *C. osculatum* sensu lato have been reported from Hokkaido, Japan [80].

Species of *Contracaecum* from birds appear to be phylogenetically separate from the seal species and confused taxonomically. Each of the well-known

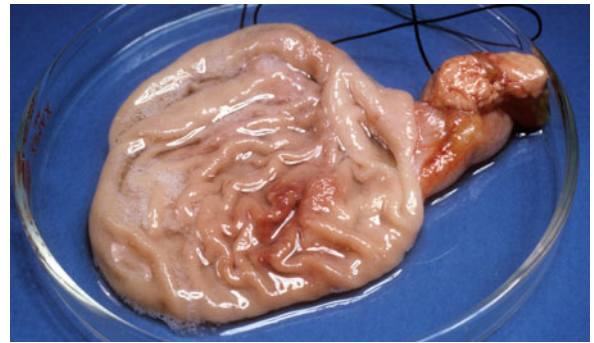


Waterborne Parasitic Diseases in Ocean. Figure 14
Commercial fillet with conspicuous specimens of *Pseudoterranova decipiens* sensu lato

species actually comprises a separate complex [81]. Members of the genus were not originally thought to infect humans, but at least two human cases of unidentified species from a bird or seal have been reported [82], and at least fourth-stage juveniles of a member of the *Contracaecum multipapillatum* complex (Fig. 15) can infect and mature in experimentally infected kittens but not rats, ducks, and chickens [83]. The worms, some attached, were in the intestine, causing hemorrhaging and associated with small ulcers.

Because members of *Hysterothylacium*, many of which were originally placed into the genus *Contracaecum*, mature in fishes, no one expected them to infect humans. Juveniles of *Hysterothylacium aduncum* have been suspected as the cause for eosinophilic granuloma in humans [84, 85], and *Hysterothylacium* type MB larva can produce pathological alterations, including eosinophilic granuloma, consistently in rodents (mice and rats) and rhesus monkey (*Macaca mulatta*) (Figs. 16 and 17) and probably humans [86]. The juvenile is small (1.4–3.1 mm long) but occurs in large numbers in some of its several secondary intermediate or paratenic fish and invertebrate hosts [87]. Worms penetrated the stomach or duodenum within an hour or so, with the attraction of an abundance of eosinophils accompanied by an apparent decrease of those leukocytes in the peripheral blood. Not all species of *Hysterothylacium* will penetrate the alimentary tract or cause lesions in mammals

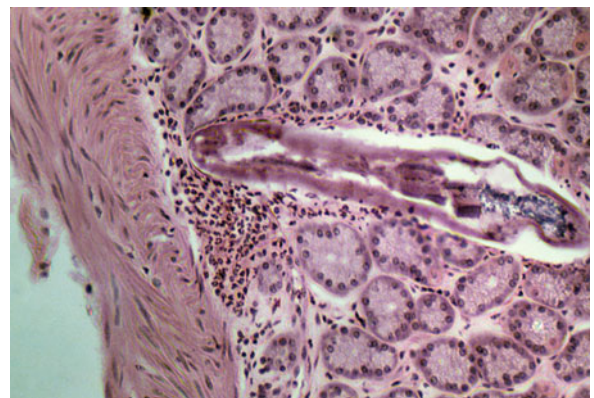
[87, 88], and juveniles of some species approach 3 cm long. A Japanese male passed a living young adult female specimen of *H. aduncum* after suffering chronic abdominal pain and diarrhea for a month. Whether the worm was acquired as a juvenile or an adult was not clear, but he had eaten raw fish, *Gadus macrocephalus* fillet pressed against kelp (Tara no kobujime) a month before onset of symptoms and *Oncorhynchus nerka* sliced and quick frozen (Ruibe), a week before onset [89]; both the cod and salmon host *H. aduncum*. The symptoms cleared following excretion of the worm. Whether *Terranova* type HA juveniles from Hawaiian



Waterborne Parasitic Diseases in Ocean. Figure 16 Stomach of rhesus monkey, *Macaca mulatta*, 5 h post-feeding with juveniles of *Hysterothylacium* type MB showing hemorrhagic lesions where juveniles penetrated



Waterborne Parasitic Diseases in Ocean. Figure 15 Red drum, *Sciaenops ocellatus*, exhibiting infective juveniles of *Contracaecum multipapillatum* sensu lato in mesentery. The same species is also common in fillets or viscera of other fish species in coastal Mississippi

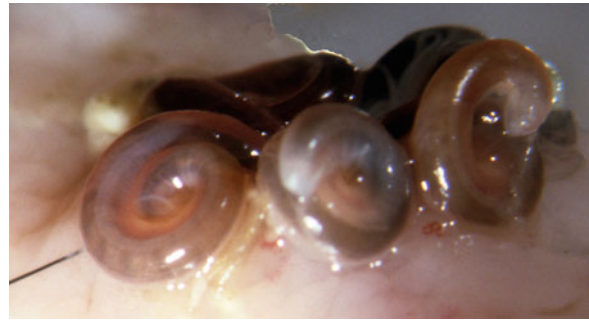


Waterborne Parasitic Diseases in Ocean. Figure 17 Section showing juvenile of *Hysterothylacium* type MB penetrating near muscularis mucosae of rhesus monkey stomach surrounded by eosinophils at 1.5 h post-feeding

bony fishes mature in sharks or in a marine mammal is not known, but, when gavaged into the laboratory rat, it migrates into the submucosa and forms a granuloma but does not pass through muscularis mucosae [90].

Other Nematodes Gnathostomatidae (Spirurina, Gnathostomatomorpha). Gnathostome nematodes are easily recognized by their spiny anterior swelling in both juveniles and adults. They are best known as public health risks of freshwater species but are mentioned here because the fishes that host the infective juveniles move into estuarine and marine habitats where they infect mammals and humans in coastal areas. *Gnathostoma spinigerum* occurs throughout Asia with infected humans primarily in Thailand and Japan, but *Gnathostoma hispidum*, *Gnathostoma doloresi*, and *Gnathostoma nipponicum* have all been identified from human cases in Japan [4, 91]. *Gnathostoma binucleatum* infects humans eating sevi- che made with estuarine fish in Sinaloa, Mexico [92], as do other species. With infections of *G. spinigerum* and *G. binucleatum*, the juveniles usually produce a migrating erythema (larval migrans) on the peripheral portions of the body for a few days to several years. Juveniles of the three other species usually migrate into surface skin, forming a serpiginous eruption on the trunk before disappearing spontaneously within 3 months. The juvenile can migrate into the central nervous system (CNS) and other vital organs such as lungs, intestine, genital organs, ear, and nose, occasionally producing a fatal disease, especially when involving the CNS [93, 94].

The typical life cycle starts with adults within a tumor-like mass in the stomach wall and serosa (Fig. 18) of the mammal host. Eggs extrude into the stomach and pass out with the feces. First-stage juveniles hatch within 12 days and develop when eaten by *Cyclops* sp. or presumably some other cyclopoid copepods. Development progresses in the copepod hemo- coel to the third-stage juvenile and are infective to a variety of fishes, amphibians, reptiles, birds, and mammals, including humans. Perhaps in some species, a specific fish serves as a second intermediate host, but fish serve as a primary paratenic host (Fig. 19), allowing the juvenile to migrate or encapsulate in any of the previously indicted paratenic hosts necessary to main- tain the population until eaten by the proper definitive



Waterborne Parasitic Diseases in Ocean. Figure 18 Tumor-like mass containing a few adult individuals of *Gnathostoma* sp. in the serosa of the stomach of the raccoon in Mississippi



Waterborne Parasitic Diseases in Ocean. Figure 19 Infective third-stage juvenile of *Gnathostoma* sp. in mesentery of the Gulf killifish, *Fundulus grandis*, in open marine lagoon of offshore barrier island, Horn Island, Mississippi

host. Human infections can come from any of the paratenic hosts and probably from drinking freshwater containing infective copepods. The importance of marine or estuarine hosts depends on the area, since infections can occur in freshwater or estuaries and freshwater hosts can migrate into estuarine or marine habitats.

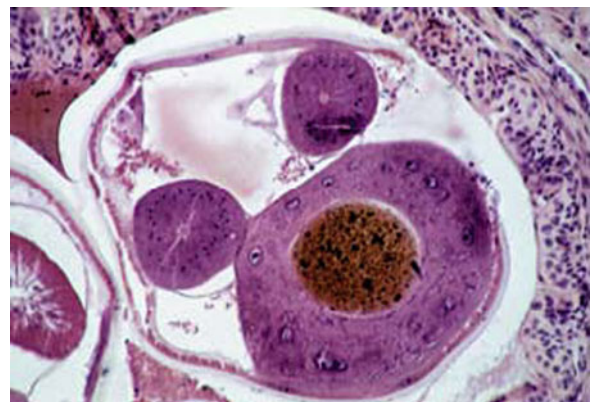
The true marine counterpart of *Gnathostoma* is *Echinocephalus*, and at least *Echinocephalus sinensis* can probably infect humans. It matures in rays and sharks in the South Pacific and is transmitted by at

least the oyster *Crassostrea gigas* and scallops. It was observed throughout the year in Hong Kong, but juveniles administered to kittens, monkeys, and puppies were infective to those mammals from August through October only [95]. When a dose of 600 worms was given to a kitten, most were apparently vomited, but the host died at 16 h. A total of 134 worms were recovered in various visceral organs. All but one of nine other kittens administered with fewer juvenile worms, 130–350, survived, but exhibited visceral larval migrans during the first 40 h. Kittens examined at days 3, 4, and 9 revealed no juvenile worm. The two different monkey species given 80 or 200 juveniles and dogs given larger doses each showed similar results of worms located throughout various visceral organs for the first 40 h post-feeding and no worm after 3 days. Substantial infections were also achieved at 18 h in kittens when fed additional worms acclimated in oysters at 28°C and 33°C [96]. Humans are potentially at risk, especially if eating heavily infected mollusks and any worms enter the CNS. There are several other species of *Echinocephalus*, and some may also have the ability to penetrate the alimentary tract of mammals and undergo a visceral or cutaneous migration. Gnathostomes cause both cutaneous and visceral migration, and human cases are probably much more common than diagnosed because specimens of the fast-moving worms are difficult to collect or biopsy; a refined serological method should be able to determine the presence of gnathostomes, and specific diagnoses should allow for a better understanding of marine associations.

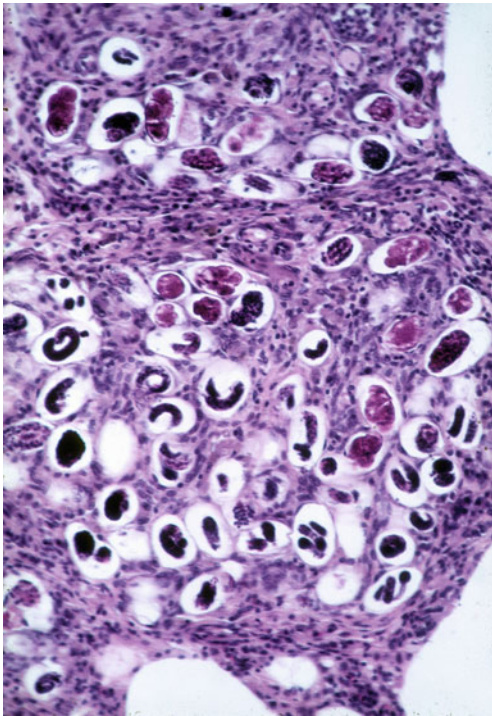
A related unidentified juvenile spiruroid nematode known as suborder Spirurina type X has been known to cause creeping eruption in at least 28 Japanese [97]. The juvenile is acquired by eating raw tiny squid (*Watasenia scintillans*) as sashimi in Japan; the serpiginous erythematous eruption along the abdomen took 4 weeks to become apparent. All infections were confirmed with an indirect immunofluorescence test.

Metastrongyloidea (Rhabditina, Strongyloidea). *Angiostrongylus cantonensis* is usually considered a freshwater or terrestrial disease, even though it was originally thought to be acquired from improperly prepared marine fishes. There are several ways that this rat metastrongyle nematode can infect humans that involve estuarine or fully marine situations, and

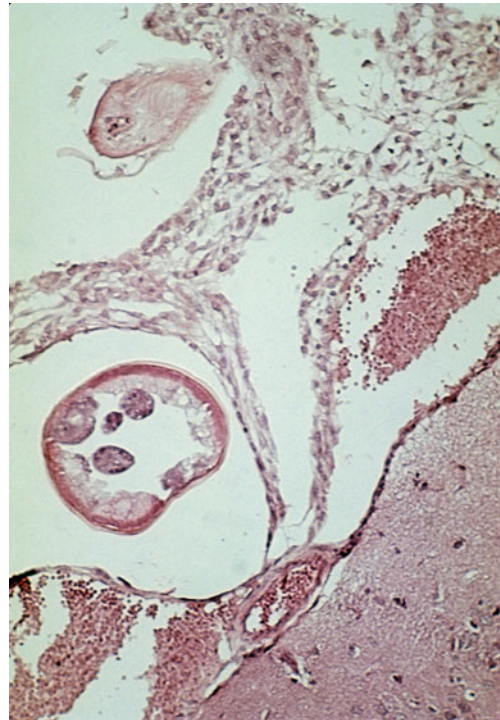
they have been reviewed [98]. The species matures in the pulmonary arteries of rats (Fig. 20) and has a rather complicated life cycle. Eggs lodge in the arteries and capillaries of the rat lung where they hatch (Fig. 21); the first-stage juvenile migrates through the alveoli and up the trachea before being swallowed and ultimately passed out in the feces. The juvenile develops to a third stage in a molluscan intermediate host. When the juvenile in the mollusk, its slime trail, or a paratenic host feeding on the host or the free juvenile is eaten by the rat, it migrates to the subarachnoid space adjacent to the brain (Fig. 22). After a couple of weeks, it forces its way into the venous system to be carried to the pulmonary arteries. Humans are accidental hosts, and the infective juvenile still goes into the brain cavity, but it remains there and often enters the brain tissue (Fig. 23) and causes behavior problems and even death in the patient. Because of abnormal behavior in infected humans, many never receive medical attention. Even though the infective juveniles do not necessarily survive as long in marine waters or in marine invertebrate or vertebrate hosts as freshwater counterparts, they survive long enough to transmit the infections. The eastern oyster, *Crassostrea virginica*, and northern quahog, *Mercenaria mercenaria*, can serve as intermediate hosts in 15 ppt, and large amounts of the



Waterborne Parasitic Diseases in Ocean. Figure 20 Section of adult female of *Angiostrongylus cantonensis* associated with other males and females in pulmonary artery of rat definitive host. The paired creamy-white uterine branches spiral around the hematin-filled intestine giving a diagnostic barber-pole appearance to this 20–30-mm-long worm when alive



Waterborne Parasitic Diseases in Ocean. Figure 21
Section of rat lung showing deposited eggs and hatching juveniles of *Angiostrongylus cantonensis* infective to mollusks

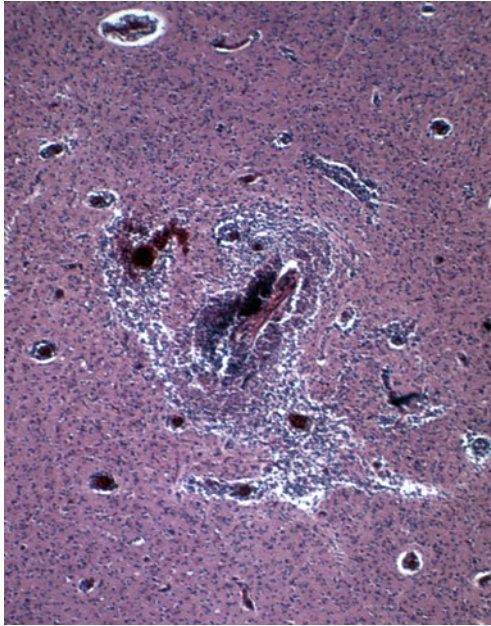


Waterborne Parasitic Diseases in Ocean. Figure 22
Section of fourth-stage juvenile of *Angiostrongylus cantonensis* developing normally in meninges of brain of rat after 29 days without any apparent inflammatory response

oyster are eaten raw. There are many possible ways for humans to acquire the infection, but it is difficult to discern individual cases. In Tahiti, special dishes involve a sauce prepared from uncooked shrimp hepatopancreas (liver/digestive gland) often containing infective juveniles mixed with grated coconut and “coconut milk.” Infections originated in the Indo-Pacific, but have spread by rats aboard ships to the Northern Hemisphere. Infections have become recognized in Hawaii, Cuba, Puerto Rico, and the USA through Louisiana.

Diectophymatidae (Dorylaimia, Diectophymatida). Juveniles of *Eustrongylides* spp., like those of *Diectophyme renale*, are relatively large and red, making ingestion seem unlikely. However, when fish intermediate or paratenic hosts are eaten alive whole, a severe peritoneal infection can occur. In the USA, eating live fish is usually a behavior to gain attention or a hazing ritual. Consequently, fishermen eating their bait, tavern clientele taking a wager, and students being initiated as

well as unaware consumers eating homemade sushi are likely victims [99]. Unlike *D. renale*, which matures in mammals, all species of *Eustrongylides* mature in birds. Three well-known species are recognized, *Eustrongylides ignotus* from North and South America plus New Zealand, usually infecting herons and egrets; *Eustrongylides tubifex* from North and South America, Europe, and into Asia in ducks, loons, and cormorants; and *Eustrongylides excisus* from Europe, Asia, and Australia in ducks, cormorants, and herons [100]. Juveniles of all should be regarded as a health risk, and all occur in both freshwater and estuaries. In Mississippi, *E. ignotus* occurs mostly in estuarine conditions and commonly spreads to its second intermediate mosquito fish and *Fundulus* hosts by a variety of tubificid and other oligochaete first intermediate hosts. Because paratenic hosts such as largemouth bass and striped bass commonly eat mosquito fish and fundulids and are preferred seafood products, humans dining on these and



Waterborne Parasitic Diseases in Ocean. Figure 23
Section of juvenile of *Angiostrongylus cantonensis* after
abnormal migration into brain of a Thai, showing atypical
inflammatory response in human brain

other fishes wild or in culture [101] in an uncooked state are at a high risk for infection.

Dracunculidae (Spirurina, Spiruromorpha, Dracunculoidea). *Dracunculus insignis*, once thought to be a synonym of *Dracunculus medinensis*, the “fiery serpent” of the Bible and possibly the “snakes” in the caduceus of the American Medical Association and many other medically related societies, constitutes a public health risk. It occurs in the hind legs of the raccoon in the brackish marshes of the northern Gulf of Mexico [1] as well as freshwater habitats throughout North America and presumably is transmitted by paratenic fish and frog hosts. Consequently, *D. insignis* could be acquired by someone eating raw fish, not presently the habit of most persons living in a region abundant in seafood typically eaten in a cooked state.

Other Parasites

Acanthocephala (Spiny-Headed Worms) Acanthocephalans, especially those that mature in the alimentary tract of marine mammals, pose a threat to those who eat raw fishes. Two palaecanthocephalans were discussed above as species that occur in the

intestine of humans, even though they did not mature in the patients. They also did not penetrate the human digestive tract. Another case of what was tentatively identified as *Bolbosoma* sp., presumably the same marine mammal parasite species mentioned earlier [63], penetrated through the intestine and became encapsulated in a granulomatous mass on the serosa of the ileum in the peritoneal cavity of a 16-year-old boy in Japan [102]. Another species, one that matures in a fish, *Acanthocephalus rauschi*, was discovered in the peritoneum of a native in Alaska. The frequency of acanthocephalan infections in humans, especially native groups from the Arctic where fishes containing infective juveniles are eaten raw, is difficult to discern but is probably substantial [18, 62].

The spiny-headed worm with hooks on the evert ing proboscis can produce severe lesions if it penetrates through the intestine or occurs in large numbers. An experimental study feeding rats with *Centrorhynchus* sp. from a paratenic host resulted in the worms penetrating the intestine on day 1, entering the muscles by day 5, and producing granulomas on day 20 [103]. When captive primates feed on intermediate hosts or are fed paratenic hosts, they can die when the worms invade beyond the serosa but otherwise produce a focal inflammatory response without harming the host [104]; when large die-offs occur, the invading worms are usually associated with secondary bacterial involvement.

Members of the Archiacanthocephala, *Moniliformis moniliformis* from rats and *Macracanthorhynchus hirudinaceus* from pig have been reported from man throughout much of the world, sometimes with serious complications. They are acquired from terrestrial arthropods such as roaches or paratenic hosts. Nine specimens of the relatively large related *Macracanthorhynchus ingens*, known from the raccoon, *Procyon lotor*, and a few other carnivores, was reported from the intestine of a 1-year-old child in Texas [105]. The worm is common in coastal areas, so it could be transmitted to humans in estuarine, freshwater, or terrestrial paratenic hosts as well as from wood roaches and millipedes, if eaten raw.

Sebekidae (Crustacea, Pentastomida, Porocephalida). Nymphs of pentastomes, a crustacean group perhaps derived from and of equal status as Branchiura (argulid parasites of fish and amphibians) also can migrate in and

become encapsulated in human visceral organs. A marine threat would be *Sebekia mississippiensis*, a parasite from the lungs of the American alligator. Nymphs are abundant in the western mosquito fish (*Gambusia affinis*) (Figs. 24 and 25) and present in the Atlantic croaker and gulf killifish, which may or may not have acquired the infection from the mosquito fish from estuarine and marine habitats, like its alligator final host. When any of the intermediate or paratenic fish, snake, or turtle hosts and possibly eggs is eaten by a mammal, the nymphs migrate and usually become encapsulated by the mammal. The opossum and river otter were infected in Mississippi, suggesting a human potential host [106], since humans are known to host nymphs of other pentastomes. This and related species infect other crocodilians. At least one case report for a related species of *Sebekia* involved skin dermatitis in a woman in Costa Rica [107]. An autopsy of a woman from Georgia, USA, with presumptive acquired immunodeficiency

syndrome (AIDS) revealed the cuticle of an unidentifiable pentastome in her pericardial sac and epicardium associated with tuberculous pericarditis [108]. Non-marine-inhabiting pentastome species of several genera present an equal risk to infect people.

Protozoans Sarcocystidae (Alveolata, Apicomplexa). The coccidian *Toxoplasma gondii* has final hosts that contain a sexual, coccidian-like phase, restricted to cats and other felines, all terrestrial mammals. Human infection results from shed oocysts in feline feces, by ingesting tissue cysts in warm-blooded mammals and birds, by eating passive vector hosts, and by congenital transmission. Infected humans, predominantly immunocompromised patients or transplant recipients on immunosuppressive therapy, can develop encephalitis, mental retardation, and blindness. If infection occurs for the first time during pregnancy, the parasite can cross the placenta into the brain or eye, and possibly cause miscarriage. The reason for the large number of human infections, perhaps a significant portion in some geographic regions, in the marine environment, may be the low number of oocysts necessary to establish a human infection. As an example, a pet cat sheds oocysts for only a few days during its lifetime, and it produces about 25 g of feces per day. However, in each fecal



Waterborne Parasitic Diseases in Ocean. Figure 24 Nymphs of *Sebekia mississippiensis* partially pulled from body cavity of intermediate host, the western mosquito fish, *Gambusia affinis*, which is infective to paratenic fish, turtle, and mammalian hosts, including humans, or to the alligator definitive host



Waterborne Parasitic Diseases in Ocean. Figure 25 Anterior end of *Sebekia mississippiensis*, a pentastome from the lung of the American alligator in Mississippi

deposit, tens of millions of oocysts may be produced, and as few as one oocyst can infect a mouse or pig [109]. The small floating oocyst can sporulate into the infectious agent at 15 or 32 ppt and remain infectious for long periods, up to 6 months at 15 ppt [110]. These oocysts, resistant to most disinfectants, however, have not been reported from coastal waters even though present in marine and coastal animals [111, 112]. When experimentally exposed to the eastern oyster, the agent in oyster tissue was infectious to mice for 6 days [113], but, even though experimentally exposed mussels (*Mytilus galloprovincialis*) tested positive by PCR assay, mice fed with various PCR-positive tissues as old as 21 days demonstrated infections for at least 3 days only when fed digestive gland tissue [114]. At least 13 marine mammals, including dolphins, have shown a positive reaction by agglutinations and indirect fluorescent antibody tests [111, 115]. The California sea otter (*Enhydra lutris nereis*) has been studied most extensively because *T. gondii* has been associated with meningoencephalitis and extensive mortalities in the California population. There are three dominant genotypes, two of which have been identified from human infections; the otter had two genotype isolates in the brain and heart, and one of those was one that was known from humans and the other a rare one [111]. The otter feeds on mollusks, and evidence indicates land-based runoff as the source of the otter infections, suggesting “potent implications for human health” [116]. Perhaps crustaceans also serve as vectors. Additional epidemiological evidence deals with Inuit women of childbearing age in Northern Quebec, Canada, where seals are infected [117]. About half were seropositive to *T. gondii*, and seropositive women were at least four times more likely to have eaten dried seal meat or seal liver than seronegative counterparts. Seronegative women during pregnancy are now instructed to avoid all uncooked or dried meat, particularly seal and caribou, and to refrain from skinning animals.

Microspora. Microspora is a large phylum with obligate intracellular members that infect and obtain their energy from primarily invertebrates and lower vertebrates, cold-blooded vertebrates. The life cycles of microsporidians include several patterns including a direct cycle, a cycle that needs a true intermediate host, a transovarian cycle, and a cycle that passes

through a vector without undergoing any development. Members of few genera infect mammals, with *Encephalitozoon cuniculi* being the best known. A few species have been reported from humans, and most of those represent immunologically incompetent individuals [118]. These include *Encephalitozoon intestinalis* and *Encephalitozoon hellem*. Others, such as *Pleistophora ronneaei*, *Trachipleistophora hominis*, and *Anncaliia vesicularium*, involve species with fish or other cold-blooded hosts. These species are named for material taken from human AIDS or otherwise immunologically incompetent patients, and they may have already been named earlier from a cold-blooded host, even a marine one. *Pleistophora ronneaei* was suspected to have been injected into a prisoner in Florida with a syringe containing infective material from a fish [1, 118]. A study conducted on other microsporidians to determine human risk from spores reaching coastal and marine waters tested the viability of the three indicated species of *Encephalitozoon* [119]. As expected for spores of species that can infect humans, the viability decreased as temperature increased from 10°C to 20°C. More proliferation of the parasites tested at 0, 10, 20, and 30 ppt occurred at the lower salinity concentrations, with *E. hellem* and *E. intestinalis* surviving better than *E. cuniculi*, in which survival lasted 1–2 weeks at 10 ppt. Those former two still proliferated at 4 weeks, with some proliferation of *E. intestinalis* at 10 ppt at 12 weeks. Consequently, all three species could infect humans and marine mammals or contaminate shellfishes. At least *E. hellem* and *E. intestinalis* can contaminate water with spores passed by aquatic birds [120]. *Enterocytozoon bieneusi* may be the most common microsporean of man. A total of 81 different genotypes with 111 different genotype names have been identified, with 26 exclusively in humans, but with none from marine animals [121]. The species without a characterized genotype was identified with PCR from an ill bottlenose dolphin (*Tursiops truncatus*) [122].

Myxozoa. Myxozoans, actually multicellular parasites that get placed in most textbooks as protozoans, seem to be bilaterian metazoans, possibly in a sister group to the Nematoda [123], or a trachylinan Cnidaria; members are not thought to infect humans. Spores occur primarily in fishes and amphibians, but atypical species have been reported from a variety of

other hosts. Sexual development takes place in oligochaetes and a variety of other invertebrates. Because people often eat fish infected by myxosporidians, they can exhibit spores from the product passed in their feces. There is no indication that any myxosporan has actually infected healthy humans and undergone vegetative reproduction. Two patients were reported to exhibit spores and diarrhea, but at least one of those was an immunosuppressed, HIV-positive patient infected with the human coccidian *Isospora belli* and the other had human helminths *Strongyloides stercoralis*, *Hymenolepis nana*, and *Ascaris lumbricoides* [124]. Other case reports, such as some in Australia involving eating fish that had been frozen, exist but probably do not represent human infections [125]. Spores of *Henneguya salminicola* also have been reported from human diarrheic feces [126, 127]. Patients had been eating salmon, and, in one case, the spores were misdiagnosed as human sperm, resulting in scandalous consequences [126].

Parasites Acquired When People Have Direct or Indirect Contact with Them

Parasites acquired by contact consist of both helminths and protozoans, but helminths usually actively enter humans by direct contact. In some cases, difficulty exists distinguishing between active and passive contact. Also diseases acquired from agents in marine water fit into different groups. Some originate from estuaries and marine habitats, but most seem to have a freshwater origin but can tolerate marine or estuarine conditions.

Direct (Active) Contact with Disease-Causing Agents

Trematoda Schistosomatidae. Members of the Schistosomatidae exhibit atypical features among trematodes such as having separate male and female individuals rather than being hermaphroditic and living in blood. They also do not have a distinctive second intermediate host, but there occurs a juvenile stage, schistosomula, that develops from the cercaria after it penetrates the definitive host. Those species that mature in humans use as their first intermediate host freshwater snails only, but some members in marine snails occurring throughout much of the world that infect birds and mammals produce cercariae that accidentally penetrate humans and establish a hypersensitivity dermatitis.

Dermatitis, known as “swimmer’s itch,” “clam digger’s itch,” and other terms, results when humans are challenged by epithelial penetration of the cercariae of some specific schistosome species occurring in the mesenteric blood vessels of the avian and mammalian marine and freshwater hosts. Swimmer’s itch is best known for freshwater species; however, several marine species result in similar human responses. Some freshwater cercariae can survive well in low-salinity conditions, but salt usually kills the snail hosts. For example, the cercariae of *Bolbophorus damnificus* can tolerate 2.5 ppt sodium chloride, a concentration toxic for the freshwater snail host *Planorbella trivolvis* that sheds it [128].

For cases of most dermatitis caused by schistosome cercariae, the cercaria hangs in or near the surface of the water where it comes into contact with the appropriate bird or mammalian (such as raccoon) host. Consequently, a pruritic maculopapular rash occurs on the human body in contact with the surface, often around the knees or ankles for those occurring in shallow water. Typically, when the cercaria penetrates human skin that it had previously penetrated with no or limited discernment by the patient, it produces a more apparent prickling sensation for several minutes. Depending on the number of cercariae and the host immune system, this condition may be associated with “macular eruption” (color change in 5–10 mm focal area without elevation or depression) with diffuse erythema or urticaria. Later in the day an itchy “macropapular eruption” (elevated colored area, 5–10 mm) develops and the papules “progress” to fluid-filled vesicles that may become secondarily infected because of the tendency to scratch them. Over the next week to month, the lesions become pigmented and apparent. The “hypersensitivity” reaction is not a true allergic one, and, when the same cercarial species or one with cross-reactivity penetrates the skin in the future, it produces a more severe dermatitis. Occasionally, systemic signs such as fever lymphadenopathy and edema develop [129].

A large number of marine schistosome species probably produce human dermatitis. The number of reported species, however, has remained low because of the inability until now with molecular tools to identify the species, the paucity of studies associating specific species with dermatitis, and the lower likelihood of human contact with cercariae in the marine

environment when compared with infective species in freshwater. Swimmers in several freshwater lakes are more prone to continual visits to cercarial/snail-laden portions of the lake. Moreover, the fact that a variety of agents in marine habitats can produce a rash, apparently including non-schistosome cercariae, makes diagnosis of the cause of the rash difficult.

Austrobilharzia variglandis is probably the best known marine cause of the dermatitis, but what is identified as *A. variglandis* probably consists of a complex of several species from California [130], Hawaii [131], Australia [129], and elsewhere as well as the clam beds of Rhode Island [132] and Connecticut [133] where it is transmitted by *Ilyanassa obsoleta* to a variety of gulls plus the double-crested cormorant and Canada goose. At least *Austrobilharzia terrigalensis* in the silver gull *Larus novaehollandiae* and *Batillaria australis* (as *Pyrasus australis*) in South Australia and Western Australia [134] has been distinguished from misidentified Australian *A. variglandis*. The snail *I. obsoleta* ranges along the Atlantic coast to Florida. *Austrobilharzia variglandis* was reported from the migratory American white pelican in Mississippi [135]. A second species from the northern Gulf of Mexico, *Austrobilharzia penneri*, occurs in *Cerithidea scalariformis*. Gulls are important definitive hosts of marine schistosomes, especially species of *Austrobilharzia*, and since gulls and shorebirds commonly occur in close association with humans, there is a tendency for infections in Hawaii, Australia, and China [129, 136]. There is need to accurately differentiate species worldwide because most or all produce dermatitis.

Members of other blood fluke genera also produce dermatitis. *Gigantobilharzia huttoni* in both the brown pelican and American white pelican and the Antilles glassy-bubble, *Haminoea antillarum* (Cephalaspidea), occurs in Florida [135, 137, 138]; *Gigantobilharzia acotylea* caused dermatitis in the Venice Canal, Italy [139]; *Ornithobilharzia* cf. *canalicula* infects terns, and *Batillaria minima* from the Gulf of Mexico in Florida, USA [140, 141], although the dermatitis attributed to this cercariae was rightfully questioned [138]. Unidentified species producing dermatitis were reported from southern California in *Littorina keehae* (as *L. planaxis*) [142], from eastern Australia in the limpet *Siphonaria denticulate* on rocky shores [143],

and from other regions such as Japan [144]. One from California occurs in the Japanese bubble snail (*Haminoea japonica*), which was introduced into San Francisco Bay in about 1999 [136]. Comparative sequence data did not allow information on whether the blood fluke was native to California, introduced from Asia in the snail, or a species from Asia acquired from migrating birds.

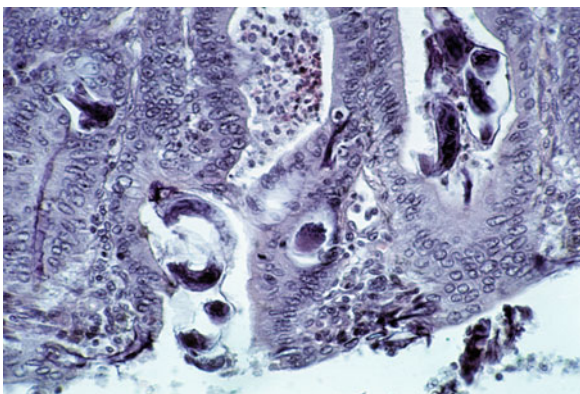
As indicated above, most schistosome species responsible for dermatitis are shed from freshwater snails in freshwater. Ten genera contain these species and are cited [145] compared with just three genera from salt or brackish waters. Studies on species of *Trichobilharzia* have provided some information that may relate to human infections. When bird species are experimentally exposed to a mouse, rather than mammalian skin trapping and eliminating the cercariae, it allows some to produce schistosomula. In the case of *Trichobilharzia regenti*, the agent typically penetrates the skin of the bird host and migrates to and becomes established in nasal tissues after passage through the peripheral and central nervous system [146]. When in a mouse, even a small number of the agent in the CNS can produce severe responses [147]. Members of *Trichobilharzia* as well as of *Gigantobilharzia* and *Ornithobilharzia*, especially on the initial exposure to experimental mammals, can progress to the lungs and CNS where they can produce severe pathological alterations [145].

Nematoda Strongyloididae (Rhabditina, Rhabditoidea). Some members of the nematode genus *Strongyloides* cause a creeping eruption (larval migrans) or rash in people exposed to juveniles and then challenged at a later date. Two species proven to cause the dermatitis in man are *Strongyloides myopotami* from the nutria (*Myocastor coypus*) and *Strongyloides procyonis* from the raccoon (*Procyon lotor*). These are probably the principal agents of what trappers, hunters, fishermen, and oil workers in the northern Gulf of Mexico call “marsh itch.” A human volunteer was exposed in his arm to a relatively small number of third-stage juveniles of *S. myopotami* on seven occasions over 11 months before a minor rash (pruritus with small papules) erupted and lasted about 5 days [148]. Following that time, the volunteer was sensitized and juveniles of both species produced a reaction.

Those from the raccoon allowed to penetrate the opposite arm produced more extensive reaction at the site of penetration for a few days, but, after disappearing for several days, it reappeared 5 cm from the site. Future exposures responded similarly but the recurrence resulted in a sinuous creeping eruption that spread from the arm to the chest. The juveniles migrating near the surface differs from the reaction by the larvae of blood flukes mentioned earlier. The species from the raccoon is morphologically similar to the human species, *Strongyloides stercoralis* (Fig. 26), which can undergo autoinfection. When a challenge dose of 185 juveniles of the raccoon species were exposed, at least one was able to mature and produce juveniles for a few days. The normal life cycle of species of *Strongyloides* differs from that of most nematodes. The threadlike parasitic female in the mucosa of the intestine deposits eggs passed in the feces as either eggs or first-stage juveniles. In the external environment, the juvenile develops into a feeding rhabditiform juvenile, which in turn develops into an infective third-stage filariform juvenile. The non-feeding filariform can invade the definitive host or develop further into a single free-living generation of both males and females. These mate and produce infective juveniles, even though under optimal conditions, the offspring may develop into a second free-living generation. Infective juveniles penetrate the skin and pass by the cutaneous blood vessels to the lungs, where they break out of the alveoli and migrate by the

respiratory tree to the pharynx and intestine, maturing in mucosal epithelium in about 2 weeks. The parasitic female produces juveniles without being fertilized and is parthenogenic. Most species of *Strongyloides* infect one or few related amphibian, reptile, mammal, and bird hosts. Probably few species would be capable of producing creeping eruption in humans, but at least two from mammals in the Southern US do. Unpublished research in Mississippi showed that the free-living infective filariform juveniles of *S. myopotami* and *S. procyonis* can tolerate seawater well.

Miscellaneous agents. Unusual circumstances can result in involvement by a parasite not expected to affect humans. For example, a diving accident resulted in such a case with an adult philometrid (Dracunculoidea) nematode, a parasite that matures in non-intestinal sites in fishes. A scuba diver in Hawaii accidentally stabbed himself when reloading a spear gun and then later cleaned a fish to eat. He noticed some parasites and discarded the fish, but 3 h later he sensed a 1-cm portion of a red worm protruding from the stab wound. He tried to kill and remove this firmly still-attached worm with vinegar, rubbing alcohol, and forceps. Because of the pain, he had the worm surgically removed by a physician [149]. Other parasites or free-living marine invertebrates can also take advantage of wounds or create their own. Isopods (Crustacea, Peracarida) in the Aegidae, Cymothoidae, Cirolanidae, and other flabelliferan families can produce lesions and urticaria, and are especially annoying. Even an amphipod has produced acute urticaria in a diver in Hawaii [1].



Waterborne Parasitic Diseases in Ocean. Figure 26 Juveniles of *Strongyloides stercoralis* passing from human intestinal crypts and representing zoonotic species, some released as eggs and others as hatched juveniles

Protozoans Acanthamoebidae (Amoebozoa). A few different free-living amoebae can cause severe disease if they can enter a human. Best known is *Naegleria fowleri* (Excavata, Heterolobosea, Vahlkampfiidae), which causes fatal “primary amoebic meningoencephalitis (PAM),” but it is restricted to freshwater habitats and moist soil throughout the world. Several species of *Acanthamoeba*, however, occur in marine waters or along marine beaches as well as in freshwater habitats. Species are known to cause chronic granulomatous amoebic encephalitis (GAE), granulomatous skin and lung lesions, and amoebic keratitis. The life cycle of *N. fowleri* consists of an amoeboid, feeding, and replicating trophozoite. During periods of low nutrients or desiccation, the motile amoeboid stage forms either

a biflagellated, non-feeding, non-replicating trophozoite or a cyst until conditions improve. Human tissues contain only the amoeboid stage. The trophozoite apparently enters through the nasal cavity and migrates into the brain and central nervous system (CNS) along the olfactory nerve. The PAM typically progresses rapidly from a sudden headache and fever to coma and death, without much possibility of diagnosis before death. Species of *Acanthamoeba* do not have a flagellated trophozoite, the amoeboid trophozoite has spike-like pseudopodia and is more sluggish, and three-layered cysts can occur in human tissues. The portal of entry may differ for different species. Invasion appears to be from a primary focus in the skin, lower respiratory tract, or nasopharynx, with dissemination of the trophozoite reaching the CNS through the circulatory system and possibly by crawling along the outside of blood vessels [150]. Signs of GAE progress much slower and initially are less severe than for PAM; they take weeks or months to result in death, and infections are difficult to diagnose. In contrast with PAM, which affects healthy people, GAE is usually associated with chronically ill, immunocompromised, or otherwise debilitated patients. Predisposing factors for infections include broad-spectrum antibiotics, steroid or antineoplastic therapy, radiation therapy, alcoholism, pregnancy, and bone marrow or renal transplantation. Many of the earlier diagnosed infections, especially in non-compromised patients, were associated with *Balamuthis mandrillaris*, a morphologically similar free-living but genetically different species that has been considered a soil-inhabiting species.

Which species of *Acanthamoeba* infect humans in the marine environment and how they infect the people have not been established. Several isolates of the *Acanthamoeba* complex have been cultured from marine waters or beaches [151–153], and eight strains are pathogenic to mice and potentially to humans. Of 17 species of free-living amoeba cultured from salt water in Northwest Spain, only *Acanthamoeba polyphaga* was pathogenic to mice [154]. Seven species have been associated with human disease, and of these, *Acanthamoeba castellanii*, *Acanthamoeba culbertsoni*, *Acanthamoeba hatchetti*, *Acanthamoeba rhysodes*, *A. polyphaga*, and *Acanthamoeba griffini* have been identified from amoebic keratitis, a hard-to-treat disease of the cornea [155]. This disease causes severe

ocular pain, photophobia, recurrent epithelial breakdown, and edema, and confirmed by trophozoites and cysts in a corneal scrape or biopsy. It was recognized in 1973 and not common until 1985 when contact lenses became popular. Lenses are maintained in a saline or tap water solution in a reusable storage case that can get contaminated.

Members of *Acanthamoeba* also have the ability to harbor viruses and bacteria, perhaps serving as vectors or reservoir hosts for a variety of agents. Isolates of *Acanthamoeba* from water sources in the Canary Islands revealed four different serotypes of adenoviruses, one related to ocular disease and the others unknown, probably respiratory, gastroenteritis, or neurologic [156]. A large DNA mimivirus was identified from *A. polyphaga*. Experimental studies [157] have shown that coxsackie B3 viruses can be adsorbed on the surface and accumulated within *A. castellanii*. The same marine amoeba can also host the bacteria *Chlamydia pneumoniae* and *Legionella pneumophila*. It did not prey on *Vibrio parahaemolyticus*, but it secreted a factor that promoted survival of the bacterium in coculture, suggesting the amoeba provides a survival to the extracellular pathogen in the environment [158]. These and other facultative parasites probably contribute greatly to transmission of microbial disease agents as can obligate parasites [159].

Nonparasitic: But Confused with Parasites In the marine environment, there are, in addition to schistosome cercariae, a variety of nonparasitic agents that cause similar appearing rashes. The stinging nematocysts of several members of the phylum Cnidaria (previously called Coelenterata) have been held responsible for “seabather’s rash” or “seabather’s eruption.” Members usually go through an alteration of generations with asexual polypoidy and sexual medusoid generations, but there are variations of this life cycle. Divers can brush against the nematocysts of fire corals (Hydrozoa, Anthomedusae, *Millepora* spp.) with their encrusting calcareous coral-like skeletons, while bathers can become entwined in expandable fishing tentacles lined by batteries of nematocysts of the “man-of-war” (Hydrozoa, Siphonophora, *Physalia* spp.) hanging from the gastrozoid, maintained on the surface by a gas-filled float. This floating colony appears like a medusa but is actually a highly modified polyp.

When the long, up to 13 m, tentacle of the Atlantic *Physalia physalis* washes onto sandy beaches, it gets ground up in the surf, leaving minute pieces with undischarged batteries of nematocysts to get into bathing suits or work gear. Medusa of several jellyfishes (Scyphozoa) have been associated with stings and rashes. Moreover, some sea wasps and box jellyfishes (Cubozoa), exhibiting square cross sections of their medusa, have been shown to cause human fatalities. The planula larvae of true jellyfishes, such as the thimble jelly, *Linuche unguiculata*, get under swimsuits, and challenge exposures produce a hypersensitive rash reaction (Fig. 27). Cases are common along the South-eastern Atlantic US, Mexico, and the Caribbean Sea, but exact identifications and prevalence of these and other cases worldwide require study. Systemic symptoms from the planulae are considered more common than with schistosome dermatitis [129]. The sting from the planula larva of the burrowing anemone (e.g., Anthozoa, *Edwardsia lineata* as *E. leidyi*), appearing as a “tiny pink egg” [160] attached to the inner lining of bathing suits, has been associated with seabather’s eruption in Long Island, New York, from mid-August through early September. The planula also parasitizes ctenophores, which can get caught in swimming apparel and transmit the condition. Not all rashes

attributed to blood flukes or cnidarians come from those agents. Many cases also result not from ejected toxins but from rubbing between the bathing suits and human skin of molluscan opisthobranch gastropod shells, sponge spicules, and spines of zoea and megalops larvae of crabs and shrimps, swarms of which can occur seasonally in heavy concentrations. Rashes attributed to “sea lice” are misleading; the term “sea lice” is a common name usually attributed to a variety of parasitic copepods but primarily *Lepeophtheirus salmonis* infesting salmon. None of these copepods produces human rashes, but some marine dermatitis caused by cnidarians has been referred to as caused by sea lice. Some polychaetes (Annelida) produce an irritating rash as does contact with the excrement of some sea cucumbers (Holothuroidea) or eating some inadequately cooked species. Contact with Cyanobacteria (previously referred to as blue-green algae), such as *Lyngbya majuscula* growing on sea grasses in Hawaiian waters, has been identified as causing “swimmer’s itch” or “seaweed dermatitis.” The agent is called fireweed in Australia. Rashes also result from some “harmful algal blooms” (red tide) throughout the world. The Internet includes some incorrect or misleading information, but researchers are tackling some of the cases, especially since they have a large impact on recreational and occupational activities in the marine and coastal environments.



Waterborne Parasitic Diseases in Ocean. Figure 27 Visiting, teaching marine biologist at San Salvador, The Bahamas, exhibiting back with hypersensitive rash developed after and presumably caused by exposure to challenge dose of planula larvae of the thimble jelly, *Linuche unguiculata*

Indirect (Passive) Contact with Disease-Causing Agents

Eating products contaminated with an agent or drinking contaminated water are herein considered differently than eating infected products because the food or water contaminated with the organism serves as a mechanical or passive vector of the agent rather than as a host that the organism requires for its survival. Also, the agent does not actively penetrate the skin as achieved by an agent of swimmer’s itch or marsh itch.

Trematodes Humans occasionally get infected with trematodes in a passive process. A person does not knowingly eat a product, and the parasite does not actively penetrate a host. For example, a seafood handler who came in contact with an abundance of fresh

killed coho salmon (*Oncorhynchus kisutch*) known to be heavily infected with the trematode *Nanophyetus salmincola* developed an infection exhibiting acute clinical signs. He denied eating raw or cold-smoked fish and rarely ate any seafood products. Alternatively, he could have transferred the minute metacercariae on his hand to his mouth while smoking [1, 43].

Protozoans Giardiinae (Excavata, Diplomonadida). Species of the flagellate *Giardia* have long been associated with severe diarrheal disease in humans. They are typically considered as freshwater or terrestrial species acquired when people drink contaminated water, but they have recently been found to be abundant in marine waters. The agents were thought to be confined to those discharged in raw sewage disposal and from recreational swimmers throughout the world, both raising public health risks. Consequently, infectious cysts or the trophozoites of the flagellates have been detected along some marine beaches near sewage outfall and canals influenced by runoff, where bathing is common in Hawaii [161], as well as in Panama and Hong Kong [111]. The trophozoite is a microscopic pear-shaped organism with four pairs of flagella. Two of these occur within the resistant cyst, and when the vertebrate host acquires the cyst, it excysts, releasing the trophozoites which reproduce asexually by binary fission and remain in the lumen or attach to the mucosa of the small intestine. Under the appropriate conditions, encystment occurs and both the cysts and trophozoites pass out with the feces. The taxonomic classification of members in the genus is not well established, but there are five recognized species (e.g., [162]), and the genetic characterization of *Giardia duodenalis*, also referred to as *Giardia intestinalis* or *Giardia lamblia* and the species with the widest host-specificity, groups at least seven genetic assemblages (genotypes) including Assemblage A, the one most commonly found in humans [163]. *Giardia* sp. occurs widespread in the river otter (*Lontra canadensis*) from the marine waters of Puget Sound Georgia Basin of Washington [164], but species also occur abundantly in the feces of a variety of temperate and Arctic marine mammals such as the right whale (*Eubalaena glacialis*, 71% prevalence), bowhead whale (*Balaena mysticetus*), ringed seal (*Phoca hispida*, 65%), California sea lion (*Zalophus californianus*), and harbor

seal (*Phoca vitulina*) but not the beluga whale (*Delphinapterus leucas*) or bearded seal. Assemblage A occurred in the harp seal (*Phoca groenlandica*), gray seal (*Halichoerus grypus*), and harbor seal (*Phoca vitulina*) but not the beluga whale [165]. Moreover, despite the common occurrence of infections in adult seals and cyst contamination on ice floes, cysts were not encountered in seal pups less than 1 year of age. However, the epidemiology and potential for zoonotic transmission in marine systems remains virtually unknown [164], but the genotypes of *G. duodenalis* in the harbor seal in Puget Sound, Washington, were the same as from a canine, from other sites, and were mostly novel genotypes [166]. Cysts remain infective for about 56 days in relatively deep freshwater cool lake water, but at least those of *Giardia muris* become inactivated by the combination of salinity and sunlight in warm Hawaiian marine beach water within 3–6 h [167]. In the Netherlands where the salinity remains stable at 28–31 ppt, *Giardia* cysts were detected in what was considered to be a low estimate of 12% of oysters, where viability was presumed to be preserved, representing some public health risk to those consumers who ate the oysters raw [168]. Cysts have also been filtered out of the water and concentrated by the blue mussel, *Mytilus edulis*, and other bivalves, and this tool has been used to monitor *G. lamblia* in Ireland [169] and for that and other species or genotypes elsewhere [112, 170]. Depending on the size and species of bivalve, it can filter as much as 20–100 l of water per day. Much information on human health risks from beaches and marine waters, animal health, source of agents, dispersal of agents, and primary vectors can be established by a variety of surveillance programs. For example, a survey of live and dead, stranded and bycatch, birds, dolphins, seals, and fishes was conducted in the Northwest Atlantic using PCR techniques [171]. With the exception of infections in seals, those of *Giardia* spp. were more frequent than those of *Cryptosporidium* spp., to be discussed next.

Coccidiasina (Chromalveolata, [Alveolata, Apicomplexan]). Coccidian species in the genus *Cryptosporidium* pose a stronger human risk for infections than those in *Giardia* [112], especially species in estuaries compared with freshwater [169]. Typically, marine mammals, invertebrates, and marine water are examined for the agents concurrently with *Giardia*

because the agents are transmitted by the fecal-oral route from humans and other mammals. Thick-walled oocysts are passed in the feces as a resting cyst, each with four sporozoites infective to the vertebrate host, where development occurs. Additionally, there is a thin-walled cyst that cannot survive long in a harsh environment, but it may lead to endogenous autoinfection. This seems to be especially true for persistent infections in immunocompromised patients. A total of 14–15 species are accepted [172, 173], and several are known to infect humans: *Cryptosporidium hominis* (previously *Cryptosporidium parvum* genotype 1), *C. parvum*, *Cryptosporidium meleagridis*, *Cryptosporidium felis*, *Cryptosporidium canis*, and *Cryptosporidium muris*, with most human infections from *C. hominis* and *C. parvum*. Human agents have been detected in more marine recreational areas than *Giardia*. They occurred in Hawaii, Panama, Puerto Rico, Australia (tidal area of Georges River), and Hong Kong [111]. In fresh deionized water, oocysts of *C. parvum* can survive for at least 6 months at 0–20°C, 3 months at 25–30°C, 1 week at 35°C [172]. As many as 10^9 to 10^{10} can be excreted from bovine calves or non-compromised humans per week, and as few as 30 oocysts have been known to infect a human volunteer [174]. Bioassays in mice show that oocysts of *C. parvum* can survive 12 weeks in 30 ppt at 10°C and at 20°C for 12 at 10 ppt, 8 weeks at 20 ppt, and 4 weeks at 30 ppt, which was long enough for the oocysts to be removed by filter feeding bivalves such as the eastern oyster (*Crassostrea virginica*) [170]. Literature reports consist of a variety of bivalve mollusks from seven countries that contained *Cryptosporidium* as detected by immunofluorescence microscopy, bioassay in mice, or molecular methods [111]. In North America, oocysts have been detected from New Brunswick to Texas. Gill washings from the oyster near a large cattle farm and septic tanks were positive for the oocysts and infective to mice [170]. The agents occur in hematocytes as well as on gill surfaces, and they were more abundant after rainfall events, with two of the four identified species (*C. hominis* and *C. parvum*) infective to humans [111]. The same two species plus *C. meleagridis* also occurred in North American commercial products [111]. Species have also accumulated in hard clams, bent mussels, and zebra mussels [175]. To date, no human cryptosporidiosis cases have been

linked to eating raw shellfish [111], even though some human genotypes have been concentrated in various bivalves. There is a higher prevalence of infection of both *Cryptosporidium* spp. and *Giardia* spp. in the ringed seal and right whale than in reported terrestrial mammals, and those infections might result from ingestion of contaminated water or from prey that had concentrated oocysts [176]. Other hosts of *Cryptosporidium* spp. include the California sea lion, bowhead whale, dugong (*Dugong dugon* in Australia), and marine foraging river otter but not the bearded seal and beluga whale [164, 176]. Isolates from the ringed seal from Nunavik, Quebec, were genetically characterized using two gene fragments and actin loci as *C. muris* plus two novel genotypes that were distantly related [177]. The risk for recreational bathers to come into contact with *C. parvum* was significantly greater on weekends than on weekdays, when the number of bathers was lower. The same was true in that Chesapeake Bay marine beach for *G. duodenalis* and *Enterocytozoon bienersi*, a microsporidian to be discussed below [178]. The proportion of water containing the agents also correlated significantly with enterococci counts. The study suggested using those counts to indicate the presence of the parasites and recommended preventing diapered children from entering the water, restricting the number of bathers in the recreational areas, advising those with gastroenteritis to avoid bathing, and using showers prior to and after bathing.

Birds contract cryptosporidiosis with at least five avian species and numerous avian genotypes, and some of these have been investigated in regard to infecting mammals. They can also serve as vectors of *Cryptosporidium* spp., *Giardia* spp., and microsporidians infective to humans. In fact, aquatic birds seem to play a substantial role in contaminating water around the world, and managing water resources could benefit by incorporating protection measures for pathogens linked to these birds [120, 179]. In contrast, little research has been conducted to evaluate the public health threat of species of *Cryptosporidium* and related genera described from marine and freshwater fishes and from amphibians and reptiles. The various taxa and isolates have been reviewed [180].

Entamoebidae (Amoebozoa). Some strains or species of amoeba *Entamoeba histolytica* complex also are

zoonotic threats. Poorly documented records of infections in the bottlenose dolphin (*Tursiops truncatus*) from Cuba, Mexico, and other locations suggest a relationship among the marine mammals, sewage, recreational beaches, and human infections. Researchers at the Gulf Coast Research Laboratory and the American Type Culture Collection have seen cysts but were unable to culture material from the bottlenose dolphin (*Tursiops truncatus*) from Mexico. Because there were slight morphological differences between it and *E. histolytica* sensu stricto, this subject deserves a critical investigation to determine if the dolphin form is restricted to marine mammals and whether it can cause fulminating dysentery, bloody diarrhea, fatigue, and abdominal pain in humans. Possibly, the agent was acquired by the dolphin from humans during associated recreational activities or from human sewage.

Parasites That Produce Excretions or Secretions That Result in an Immunological Response in People

Ascaridoid juvenile nematodes produce excretions and secretions (ES products) that elicit immunological responses. These responses include both the allergic response by the host to living worms migrating into or through tissues and true anaphylactic reactions triggered by dead or living worms by food-borne, airborne, and skin contact routes [75]. Ascaridoids are unusual among nematodes to produce such reactions. Most concerning is that some sensitized individuals respond to infected cooked seafood products and even chicken that has been fed fish meal made from infected fish products [181].

The tissue migration by ascaridoids is made possible by potent proteolytic enzymes released from the esophageal gland and excretory cell through separate anterior openings. These contain hyaluronidase, serine proteases, anticoagulants, and numerous other identified and unidentified substances, but only specific ones probably produce the well-defined, erosive, hemorrhagic lesions in the gastric mucosa and these and others produce other effects in the human accidental host. Metabolic products released by the migrating juvenile also produce humeral and cellular responses, especially involving the acute lesions. These products,

also including surface and somatic components of the juvenile, form insoluble immune responses with antibody. Some also cause direct IgE-independent degranulation of mast cells, at least in sensitized mice, and chemotaxis of eosinophils associated with thermolabile factors from the parasite, and produce the characteristic local response in the digestive tract but not systemically like in most other helminth infections [75]. When either living or dead individuals of *A. simplex* are ingested, cholinergic hyperactivity and adrenergic blockade can be rapidly induced, resulting in focal reactions associated with living worms but widespread along the entire bowel if the worm becomes ruptured or disintegrated. These activities may explain why 70% of the symptoms in *A. simplex*-induced anaphylactic reactions involve the digestive tract.

Anaphylaxis appears to be complex. Rodent studies demonstrate a mixed Th1/Th2 pattern of cytokines when the host is sensitized and challenged intravenously with certain parasites, but not when challenged orally. When the antigen is an extract of *A. simplex* proteins or a live worm, a strong Th2 response occurs, resulting in scratching, irritability, diarrhea, and puffiness around the eyes within 1 h. The multifaceted immune hypersensitivity reactions after induction by members of the *Anisakis simplex* complex have been investigated in terms of a wide range of parasite products and host cells; a concise helpful review treats all the known facets [75]. Hypersensitivity is usually diagnosed by skin prick tests and in vitro confirmation (specific IgE, histamine release, and basophil activation test). The skin prick test for *A. simplex* was first used in 1995 and now is used by physicians as an important test for cases of urticaria and anaphylaxis [75, 182, 183].

Recognition of allergic signs has been emphasized because of cases in Spain and South Africa. In Spain where fish is consumed abundantly, what was referred to as *A. simplex* is the most important hidden food allergen in the adult population suffering acute urticaria and anaphylaxis. It comprises as much as 10% of the anaphylaxis previously diagnosed as idiopathic [75], if the specific immunoglobulin E (IgE) detection by ImmunoCAP assay did not overestimate the number of sensitized subjects. Anisakiasis in Spain is caused primarily by eating pickled anchovy (*Engraulis encrasicolus*). That fish plus European hake (*Merluccius merluccius*) and cod (*Gadus morhua*,

just in Atlantic Ocean rather than in both the Mediterranean Sea and Atlantic Ocean) are eaten by allergic patients. Half the 64 allergic patients in one report required emergency treatment [184]. Half of the patients presenting signs of infection said they ingested raw fish, but the remainder ate cooked fish or, in rare cases, canned fish. Allergic cases requiring hospitalization demonstrated respiratory arrest, severe shock, and persistent angioedema.

In Spain, the matter of allergic responses in “gastroallergic anisakiasis” remains complicated and seems to be reported primarily from some regions in Spain. Patients entering the hospital emergency room with a severe acute epigastric pain and positive *Anisakis* prick tests plus total and specific IgE assays were divided into two groups. One group was diagnosed by the presence of at least one worm in the stomach, and the other group which had no worm seen in the stomach. Most patients had just eaten raw or pickled anchovies. Both groups averaged about 5 h, but up to 26 h, between intake of raw fish and onset of hypersensitivity symptoms, and no significant difference occurred between the two in their allergic symptoms of urticaria, angioedema, erythema, bronchospasm, and anaphylaxis, suggesting a borderline condition between a food allergy and parasitic disease supporting quick removal of the worm from the stomach. Moreover, only 26 of 40 patients required drug therapy to manage the allergic reaction, with only three requiring it for more than 1 day, suggesting an acute, self-limiting disease. In contrast, anaphylaxis can be life threatening, but abdominal symptoms disappeared within a few hours after removal of the worm [185]. When a portion of the *Anisakis*-sensitized groups with gastric worms, without gastric worms, and with no gastroscopy conducted (when symptoms did not persist for more than 8 h) were challenged with frozen worms from the blue whiting (*Micromesistius poutassou*, from either Atlantic or Mediterranean sources) and later from undocumented hosts on multiple occasions, there was no patient who suffered a reaction. This lack of a response to the thermostable proteins suggested that live worms had to be present for the allergic symptoms to occur [186]. Because the specimens from the Mediterranean Sea were from the anchovy, they were *A. pegreffii* [70] and not *A. simplex* as reported, even though if some material came from

the Atlantic Ocean, both species could have been present. However, human infections with *A. pegreffii* are typically gastric in nature, and those of *A. simplex* are usually intestinal, based on findings from Japan, where 10% of healthy adults were seropositive against the *Anisakis* antigen [187]. The fact that some individuals exhibit allergic episodes when no viable nematode is present may express a difference between the effects of *A. pegreffii* and some other species of *Anisakis* or may be variation in responses by individual patients.

The cases in Western Cape Province, South Africa, involved workers in fish-processing plants. An epidemiological study of 578 workers from two large-scale plants reported 30 workers (5%), who said they had allergic symptoms to seafood. A total of 87% said symptoms occurred after eating fish, 40% after handling them, and 17% after smelling them. A prevalence of sensitization was tested as 6% with a higher value of 8% sensitive to *Anisakis* [188, 189]. *Anisakis*-specific IgE reactivity in the workers was determined to be associated with bronchial hyperreactivity and dermatitis. In corresponding mice studies, the juvenile worm induced a striking Th2/type 2 response. The investigation suggested that consumers acquiring an infection most likely can acquire sensitization to anaphylaxis, but that exposure to *Anisakis* proteins alone may be enough to elicit allergic reactions in sensitized individuals. In other words, occupational exposure to infected fish or fish meal constitutes a risk factor for developing sensitization to *Anisakis* or other ascaridoid-related allergic disease.

Historically, eating fish fillets infected with the plerocercoids of trypanorhynchian cestodes has been considered harmless with rare exception. All members of the group mature in elasmobranchs. In fact, some infected intermediate host products are preferred over noninfected counterparts; one example is the flesh of the Atlantic pomfret, *Brama brama*, also known by a variety of other common names, infected with plerocercoids of *Gymnorhynchus gigas* and preferred by some Portuguese [99]. Recently, the rat and mouse were orally inoculated with plerocercoids of *G. gigas* to test for anti-worm IgG, M, and A (H + L) levels in intestinal fluids and serum as well as specific serum IgE levels by enzyme-linked immunosorbent assay (ELISA) [190]. Levels of all increased in the challenged mouse, producing distress. The rat had an increased expression of

heat shock proteins in the intestine and spleen. Repeated exposure to the worm in the rodents produced clinical signs appearing progressively more rapid and lasting longer, suggesting that feeding on infected fish triggered production of anaphylactic-type antibodies in rat, mouse, and, by implication, human.

Management, Control, and Treatment of Parasites Affecting People

Management

Management starts with education about parasites. Parasites are or may be involved with a recognized or unrecognized problem but are not necessarily harmful or bad, if certain practices are recognized or followed by regulating agencies, funding agencies, the seafood industry, consumers, and those using marine waters for occupation or recreation. Harm can be done if partial data deludes people not to eat seafood or not to enter the water.

Being alerted to public health risks allows people to make choices about what product to eat, how it should be prepared, when to avoid a product, when not to enter specific bodies of water, and when to visit a physician. If the industry provides a safe and sanitary product to its customers, the number of customers will increase.

Control

The best way for processors and restaurants to control infections with most helminths and most protozoans is to freeze or heat the product. Some restaurants flash-freeze products sold as fresh to avoid infective products, and apparently few customers can detect the difference from the never frozen products. If products with few or no parasites can be provided to these groups by fishermen or fish farmers, that product will be in higher demand, able to meet regulations, and more valuable. Controlling infections by interrupting a parasite life history link and harvesting from a noninfected region are easy ways to avoid future problems.

Heating and Freezing

Historically, the US Food and Drug Administration (FDA) [191] has recommended food-service industries

to cook fish products to an internal temperature of at least 63°C by conventional methods or to an internal temperature of 74°C by a microwave process rotated midway through the process and allowed to stand for 2 min afterward. Any process sufficient to kill bacterial pathogens also will kill parasites.

However, the FDA, under the seafood Hazard Analysis and Critical Control Point (HACCP) regulations, will allow for a facility to submit to it an alternative proposal with justification to modify the recommendations. When ten specimens of *Anisakis simplex* sensu lato were embedded into an approximate 1.8-cm-thick portion of a fillet of the arrowtooth flounder (*Atheresthes stomias*) and microwaved, the viability of the worms was assessed. Survival was determined as 31% at 60°C, 11% at 65°C, 3% at 74°C, and 0% at 77°C [192]. Thick salmon fillets with *Dipyllobothrium* spp. and *Anisakis* spp. require additional cooking. Microwaving pork chops with *Trichinella spiralis* still had active juveniles at 82°C, and microwaving can produce temperatures differing by 20°C when measured just 1 cm apart. The purpose of HACCP Regulation, as implemented in 1997, is for processors of fish and fisheries products to develop and implement reasonable plans for safe and sanitary procedures. Recommendations are established at each point that specific parasites can be detected, eliminated, or treated, and FDA will guide and work with those people and facilities in the industry to provide their specific appropriate plans and systems (www.fda.gov/Food/FoodSafety).

Freezing infected fisheries products is also not clear because the effectiveness depends on the temperature of the process, the length of time undergoing freezing, the length of time held frozen, the fat content of the product, and the species and stage of the parasite. For example, most cestode plerocercoids are more susceptible to freezing than nematode juveniles, which in turn are more susceptible than encysted trematode metacercariae. FDA recommends the following: freezing and storing at −20°C or below for 7 days, freezing at −31°C or below until solid and then storing at that temperature for 15 h, or at −4°C for 24 h. The fishing industry has used a process called “blast freezing,” which rapidly reduces the temperature to −40°C, and it has shown to effectively kill parasites and not have much influence on flavor or texture of the product. Also, most US sushi bars freeze their fish [1].

Other communities have different recommendations or regulations. The European Food Safety Authority presented a scientific opinion on food safety related to parasites in fishery products [193]. The presence of *Anisakis* was of major concern for products intended for marinating, salting, and eating fresh. Treatments should provide an equivalent level of protection as freezing at -20°C for ≥ 24 h, at -35°C for ≥ 15 h, or -15°C for ≥ 96 h at the product's core; heat treatment should be $>60^{\circ}\text{C}$ for at least 1 min. These are less rigorous than those in the USA but more than some other regulations. Emphasis focused on the insufficiency of many traditional marinating and cold-smoking methods to kill the juvenile worms. Some metacercariae can tolerate more-harsh temperatures and treatments than other parasites.

Detection of Parasites and Culling

Large parasites and those that contrast well with the infected seafood product are those most easy to detect. Many of these are not harmful to consumers, but they will have an influence on whether consumers will purchase the products or eat them in a "raw" state. Some, such as *Anisakis* spp. and *Pseudoterranova* spp., constitute a public health risk, and these have attracted research in detection methodology. The efficiency of successively more accurate methods for detecting these nematodes range from (1) gross visual inspection, (2) candling on a light table, (3) candling of belly flaps (ventral portion of fish), (4) ultraviolet illumination (UV) of frozen fillets, (5) pepsin/hydrochloric acid degradation, and (6) UV illumination of frozen remains [194]. Candling or passing a product over a light table, either sandwiched between two glass or plastic plates, sliced fillet, or whole fillet, has been a popular method for several decades and usually the recommended method during industrial processing of marine fish intended for human consumption. European Union regulation 91/493/EEC and Norwegian fish quality regulations require removal of any visible parasites seen in gross inspection of commercial products by spot checking [194]. The more efficient methods indicated above are occasionally used for scientific purposes, and in one article [194] examining Norwegian spawning herring, mackerel, and blue whiting from the Northeast Atlantic

Ocean for nematode juveniles in fillets, the detection efficiency for candling was 10% or less. After testing with the other listed methods, the highest values were obtained with UV illumination of remains after enzymatic digestion. Size and thickness of fillet as well as texture and color of flesh influence the efficiency. Pressing 2–3 mm layers of fillet with a hydraulic and manual press between 12-mm-thick acrylic glass sheets, deep-freezing the removed cake in a plastic bag, and examining the sample in the bag with UV light at 366 nm more than doubles the recovery rate of candling at 1,500 lx; this method is rapid and good for surveys [195]. Imaging spectroscopy provides a promising method to detect spectral and spatial information from nematodes in cod as deep as 0.8 cm below the surface of the fillet. That is deeper than can be determined by manual inspection, but the instrumentation has not been perfected enough for commercial use [196].

Fish Farming Practices

In Europe, apart from farmed Atlantic salmon, sufficient monitoring data on marine products are not available [193]. A wealth of information occurs in the literature and on Internet sites about how to detect and treat parasites and control infections, usually with focus on health of the cultured product rather than on a human consuming or coming into contact with the product. This chapter does not recommend or discuss the numerous specific methods for use in fish farms and other aquaculture operations, but parasites are best managed in those facilities by breaking the link in the life cycle that can disrupt or eliminate the infections. Fish in culture can have more or fewer parasites than wild products, and the degree relates to several factors [1].

Treatment

This chapter mentions a few chemotherapeutic treatments for infected patients but purposely does not attempt to provide compounds for all taxa. People are becoming resistant to some compounds, some compounds have serious side effects, the Internet contains both good and poor or misleading advice, and new compounds are continually being tested or made available. Because a reader should use or recommend the

most reliable treatment for the agent or strain, the reader should seek professional advice. Recent information has been compiled such as that found in the most recent edition by Garcia listed in the “Books and Reviews” or some other recent source. The Centers for Disease Control and Prevention (CDC) has a website with much helpful information and a willingness to respond to individuals that contact it. On the other hand, methods to avoid infections by properly freezing or cooking a product occur above. Of course, because of well-entrenched customs, the wonderful flavor of raw or specially prepared seafood, and the lack of knowledge about most parasites, inadequately prepared and uncooked products will always pose a risk of infection.

Future Directions

As seen already in this chapter, a wealth of information has been acquired in the last few decades regarding parasites that pose a public health risk. Many questions remain unanswered and lifetimes of research will be required to answer these and additional questions regarding the risks and what to do about them. Some of the suggested paths for future research occur below in five overlapping categories:

1. Harmful parasites presently unknown or unrecognized still need to be discovered. Because of new cuisines, opportunities to get infected, methods to characterize agents, and methods to detect agents as well as variability in human responses to parasites, those parasites need to be determined that are capable of infecting humans. As seafood products are overfished or the consumers increase in number, seafood products need to come from new areas or new products. Consequently, the risk for acquiring an infection from a product historically recognized as safe increases without the consumer's knowledge. Protozoans cannot be seen without a microscope, and infections by marine species and freshwater species flushed into the marine environment are being recognized as important agents affecting human health, and the marine environment probably contains many more of these than recognized.
2. An improvement in taxonomy of many parasitic groups will allow the scientific community and public to know what species or genotypes are involved with diseases in different areas resulting from different parasite life histories. Morphologically similar species can elicit entirely different responses, and the knowledge of what specific agents are where and what signs of disease result from which of those agents comprises a fundamental cornerstone of biology and research. Taxonomic research will involve morphological treatment of quality specimens from both humans and natural hosts backed up with corresponding molecular characteristics. Molecular tools (PCR, quantitative PCR, sequence analysis, and other techniques) allow better detection and diagnosis of some of these parasites and consequently improve our understanding of the epidemiology of human infections, the geographic range of the agents, the hosts of the agents, and the longevity of the parasites. Taxonomic studies matching genetic sequences with corresponding morphological data help identify or establish sister and cryptic (closely related) species or strains. This is true for helminths as well as protozoan groups as indicated in the discussion of specific taxa. In addition to needing supplemental molecular research on taxonomy, identifications, and diagnoses, molecular tools will help determine/confirm life cycles, life histories, mechanisms of pathogenesis of infections, and approaches to manage, control, and treat the specific agents.
3. Once specific agents are characterized, methods to detect and diagnose them need improvement, especially by noninvasive means. This is especially important because many parasites do not mature in humans, thereby not allowing the detection of diagnostic eggs, cysts, and spores in the feces or blood. Even when parasites do mature in patients, they cannot always be specifically identified. For example, most heterophyids have very similar eggs, and biological variation occurs for each species. Cross-reactivity occurs in many of the present serological and molecular assays. In the case of *Paracapillaria philippinensis*, there is a need to identify juvenile stages as well as serologic or other diagnostic methods to assess infections when eggs and adult specimens are not apparent in fecal examinations. Improved molecular assays could also

detect specific infected fishes in specific locations that are responsible for most of the resulting human infections. More sensitive methods are necessary to detect general and specific isolates of *Cryptosporidium*, *Giardia*, and *T. gondii* in bivalves, where some unknown and nonremovable components inhibit the presently used PCR reactions [168]. As indicated above, many more protozoans probably infect humans, and new methods will allow their detection.

4. Surveillance methods of known agents require improvement so that infections can be reduced, and causes can be detected. In Europe, the EFSA recommends coordinated studies to improve and implement surveillance and diagnostic awareness of allergic reactions to parasites in fishery products and encourages epidemiological studies throughout Europe to assess the impact of *Anisakis* on human-associated disease, including all allergic forms [193]. All surveillance programs in the marine environment are in an early stage of research, and many could be implemented in conjunction with microbiological studies involved with maintaining healthy waters.
5. There is a need to create ways to avoid parasitic infections without having to reduce seafood consumption or reduce time spent in the water for recreation or occupation. The spread of information in a subtle manner should definitely reduce parasite infections in most countries today. On the other hand, this can be difficult when an improved economy in areas such as Vietnam have increased the social activities that are conducive to eating more raw seafood because it is “healthier when eaten raw” and acquiring more parasites. Also, global climate changes can affect infections as exemplified earlier for the peoples of Hokkaido Island, Japan, who understood parasitic infections and depended on their seafood products.

A few causes for human infection by zoonotic parasites that seem correctable by means other than spreading knowledge about specific parasites include unsanitary defecation habits, use of human excreta as fertilizer, inadequate sewerage systems, and inadequate aquaculture practices. New ideas associated with acquired knowledge about infective parasites and

their life histories can contribute to modifications in sources of products and methods of complex food consumption and cooking habits. The latter include economic and sociocultural factors such as beliefs and tradition. Examples of traditional dishes in addition to those already mentioned include raw or partially cooked aquatic products such as raw crab soaked in soy sauce (ke-jang, Korea), raw drunken crabs and raw grass carp in China, and raw fish in Thailand (Koi-pla [raw fish with chopped garlic, lemon juice, chili, rice, and vegetables] and pla som). In contrast, in some industrialized countries like Japan, where eating raw fish is widespread, infections can be coupled with foreign travel and with eating imported foods or exotic delicacies.

Acknowledgments

I thank Janet Wright for help with organization of the literature and Kim Overstreet and Jean Jovonovich for reading the draft. Some of the figured histological sections are from the author's old photographs taken of slides in the collection of the late Paul C. Beaver at Tulane Medical School. The review is based on work supported by USDC, NOAA, award no. NA08NOS4730322 and subaward no. NA17FU2841, the NSF under grant no. 0529684, and US Wildlife, Fisheries, and Parks, CIAP, award M10AF2015, MS.R. 798.

Bibliography

Primary Literature

1. Deardorff TL, Overstreet RM (1991) Seafood-transmitted zoonoses in the United States: the fishes, the dishes, and the worms. In: Ward DR, Hackney CR (eds) Microbiology of marine food products. Van Nostrand Reinhold, New York, pp 211–265
2. Fischler C (2002) Food selection and risk perception. In: Anderson GH, Blundell J, Chiva M (eds) Proceedings from the symposium food selection, from genes to culture. Danone Inst., Paris, France, pp 135–151
3. Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Taylor FJR (2005) The new higher level classification of Eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol 52:399–451
4. Nawa Y, Hatz C, Blum J (2005) Sushi delights and parasites: the risk of fishborne and foodborne parasitic zoonoses in Asia. Clin Infect Dis 41:1297–1303
5. Ishikura H, Takahashi S, Yagi K, Nakamura K, Kon S, Matsuura A, Sato N, Kikuchi K (1998) Epidemiology: global aspects of

- anisakidosis. In: Tada I, Kojima S, Tsuje M (eds), ICOPA IX 9th International Congress of Parasitology, Bologna, Italy, pp 379–382
6. Kuchta R, Scholz T, Brabec J, Bray RA (2008) Suppression of the tapeworm Order Pseudophyllidea (Platyhelminthes: Eucestoda) and the proposal of two new orders, Bothriocephalidea and Diphyllbothriidea. *Int J Parasitol* 38:49–55
 7. Chung PR, Wm S, Jung Y, Pai SH, Nam MS (1997) Five human cases of *Diphyllbothrium latum* infection through eating raw flesh of redlip mullet, *Liza haematocheila*. *Korean J Parasitol* 35:283–289, Article in Korean
 8. Rausch RL, Adams AM, Margolis L (2010) Identity of *Diphyllbothrium* spp. (Cestoda: Diphyllbothriidae) from sea lions and people along the Pacific coast of South America. *J Parasitol* 96:359–365
 9. Baer KG (1969) *Diphyllbothrium pacificum*, a tapeworm from sea lions endemic in man along the coastal area of Peru. *J Fish Res Board Can* 126:717–723
 10. Adams AM, Rausch RL (1997) Diphyllbothriasis. In: Connor DH, Chandler FW, Schwartz DA, Manz HF (eds) Pathology of infectious diseases, vol II, Appleton and Lange. Stamford, Connecticut, pp 1377–1389
 11. Yamane Y, Shiwaku K (2003) *Diphyllbothrium nihonkaiense* and other marine-origin cestodes. *Prog Med Parasitol Jpn* 8:245–259, Meguro Parasitological Museum (Tokyo)
 12. Scholz T, Garcia HH, Kuchta R, Wicht B (2009) Update on the human broad tapeworm (Genus *Diphyllbothrium*), including clinical relevance. *Clin Microbiol Rev* 22:146–160
 13. Arizono N, Yamada M, Fukumoto S, Nakamura-Uchiyama F, Ohnishi K (2009) Diphyllbothriasis associated with eating raw pacific salmon. *Emerg Infect Dis* 15:866–870
 14. Wicht B, Scholz T, Peduzzi R, Kuchta R (2008) First record of human infection with the tapeworm *Diphyllbothrium nihonkaiense* in North America. *Am J Trop Med Hyg* 78: 235–238
 15. Wicht B, de Marval F, Peduzzi R (2007) *Diphyllbothrium nihonkaiense* (Yamane et al., 1986) in Switzerland: first molecular evidence and case reports. *Parasitol Int* 56: 195–199
 16. Yamasaki H, Kuramochi T (2009) A case of *Diphyllbothrium nihonkaiense* infection possibly linked to salmon consumption in New Zealand. *Parasitol Res* 105:583–586
 17. Margolis L, Rausch RL, Robertson E (1973) *Diphyllbothrium ursi* from man in British Columbia – first report of this tapeworm in Canada. *Canadian J Public Health* 64:588–589
 18. Rausch RL, Adams AM (2000) Natural transfer of helminths of marine origin to freshwater fishes, with observations on the development of *Diphyllbothrium alascense*. *J Parasitol* 86:319–327
 19. Chung DI, Kong HH, Moon CH, Choi DW, Kim TH, Lee DW, Park JJ (1995) The first human case of *Diplogonoporus balaenopterae* (Cestoda: Diphyllbothriidae) infection in Korea. *Korean J Parasitol* 33:225–230
 20. Clavel A, Bargues MD, Castillo JF, Rubio MD, Mas-Coma S (1997) Diplogonoporiasis presumably introduced into Spain: first confirmed case of human infection acquired outside the Far East. *Am J Trop Med Hyg* 57:317–320
 21. Arizono N, Fukumoto S, Tademoto S, Yamada M, Uchikawa R, Tegoshi T, Kuramochi T (2008) Diplogonoporiasis in Japan: genetic analyses of five clinical isolates. *Parasitol Int* 57: 212–216
 22. Skeríková A, Brabec J, Kuchta R, Jiménez J, García HH, Scholz T (2006) Is the human-infecting *Diphyllbothrium pacificum* a valid species or just a South American population of the holarctic fish broad tapeworm, *D. latum*? *Am J Trop Med Hyg* 75:307–310
 23. Holiday DM, Guillín S, Richardson DJ (2003) Diphyllbothriasis of the Chiribaya culture (700–1476 AD) of southern Peru. *Comp Parasitol* 70:167–171
 24. Kamo H (1981) Present situation of human diphyllbothriasis in Japan. *Yonago Acta Med* 25:144–155
 25. Chai J-Y (2007) Intestinal flukes. In: Murrell KD, Fried B (eds) Food-borne parasitic zoonoses: fish and plant-borne parasites. Springer, New York, pp 53–115
 26. Chai JY, Lee SH (2002) Food-borne intestinal trematode infections in the Republic of Korea. *Parasitol Int* 51:129–154
 27. Cho S-H, Cho P-Y, Lee D-M, Kim T-S, Kim I-S, Hwang E-J, Na B-K, Sohn W-M (2010) Epidemiological survey on the infection of intestinal flukes in residents of Muan-gun, Jeollanam-do, the Republic of Korea. *Korean J Parasitol* 48:133–138
 28. Dung DT, De NV, Waikagul J, Dalsgaard A, Chai JY, Sohn WM, Murrell KD (2007) Fishborne intestinal zoonotic trematodiasis, Vietnam. *Emerg Infect Dis* 13:1828–1833
 29. Vo DT, Murrell D, Dalsgaard A, Bristow G, Nguyen DH, Bui TN, Vo DT (2008) Prevalence of zoonotic metacercariae in two species of grouper, *Epinephelus coioides* and *Epinephelus bleekeri*, and flathead mullet, *Mugil cephalus*, in Vietnam. *Korean J Parasitol* 46:77–82
 30. Perry B, Sones K (2007) Poverty reduction through animal health. *Science* 315:333–334
 31. Yang J-J, Guk S-M, Han E-T, Chai J-Y (2000) Molecular differentiation of three species of *Metagonimus* by simple sequence repeat anchored polymerase chain reaction (SSR-PCR) amplification. *J Parasitol* 86:1170–1172
 32. Font WF, Overstreet RM, Heard RW (1984) Taxonomy and biology of *Phagicola nana* (Digenea: Heterophyidae). *Trans Am Microscop Soc* 103:408–422
 33. Overstreet RM (1978) Marine maladies? Worms, germs, and other symbionts from the northern Gulf of Mexico. Mississippi-Alabama sea grant consortium, MASGP-78-021, Blossman Printing, Ocean Springs, MS, 140p
 34. Welberry AE, Pacetti W (1954) Intestinal fluke infestation in a native Negro child. *Bull Dade County Med Assoc* 24(34):45
 35. Adams KO, Jungkind JL, Bergquist EJ, Wirts CW (1986) Intestinal fluke infection as a result of eating sushi. *Am J Clin Pathol* 86:688–689
 36. Youssef FG, Mikhail EM, Mansour NS (1989) Intestinal capillariasis in Egypt: a case report. *Am J Trop Med Hyg* 40:195–196

37. Mitchell AJ, Overstreet RM, Goodwin AE, Brandt TM (2005) Spread of an exotic fish-gill trematode: a far-reaching and complex problem. *Fisheries* 30(8):11–15
38. Africa CM, de Leon W, Garcia EY (1940) Visceral complications of intestinal heterophyidiasis of man. Monographic series no. 1, University of the Philippines, Manila
39. Kean BH, Breslau RC (1964) Cardiac heterophyidiasis. In: *Parasites of the Human Heart*. Grune and Stratton, New York, pp 95–103
40. Deschiens R, Collomb H, Demarchi J (1958) Distomastose cerebrale a *Heterophyes heterophyes*. In: Abstracts of the sixth international congress of tropical medicine and malaria. Lisbon
41. Paperna I, Overstreet RM (1981) Parasites and diseases of mullets (Mugilidae). In: Oren OH (ed) *Aquaculture of grey mullets*. Cambridge University Press, Cambridge, pp 411–493
42. Heard RW, Overstreet RM (1983) Taxonomy and life histories of two North American species of "*Carneophallus*" (= *Microphallus*) (Digenea: Microphallidae). *Proc Helminthol Soc Wash* 50:170–174
43. Eastburn RL, Fritsche TR, Terhune CA Jr (1987) Human intestinal infection with *Nanophyetus salmincola* from salmonid fishes. *Am J Trop Med Hyg* 36:586–591
44. Chai J-Y, Han E-T, Park Y-K, Guk S-M, Lee S-H (2001) *Acanthoparyphium tyosenense*: the discovery of human infection and identification of its source. *J Parasitol* 87:794–800
45. Little JW, Hopkins SH, Schlicht FG (1966) *Acanthoparyphium spinulosum* (Trematoda Echinostomatidae) in oysters at Port Isabel. *Texas J Parasitol* 52:663
46. Guk S-M, Kim J-L, Park J-H, Chai J-Y (2007) A human case of *Plagiorchis vespertilionis* (Digenea: Plagiorchiidae) infection in the Republic of Korea. *J Parasitol* 93:1225–1227
47. Justo MCN, Tortelly R, Menezes RC, Kohn A (2008) First record in South America of *Didymosulcus palati* and *Didymosulcus philbranchiarca* (Digenea, Didymozoidae) with new hosts records and pathological alterations. *Mem Inst Oswaldo Cruz* 103:207–210
48. Murrell KD, Pozio E (2000) Trichinellosis: the zoonosis that won't go quietly. *Int J Parasitol* 30:1339–1349
49. Rausch R, Babero BB, Rausch RV, Schiller EL (1956) Studies on the helminth fauna of Alaska. XXVII. The occurrence of larvae of *Trichinella spiralis* in Alaskan mammals. *J Parasitol* 42:259–271
50. Leclair D, Forbes LB, Suppa S, Proulx J-F, Gajadhar AA (2004) A preliminary investigation on the infectivity of *Trichinella* larvae in traditional preparations of walrus meat. *Parasitol Res* 93:507–509
51. Pozio E, La Rosa G, Rossi P, Fico R (1989) Survival of *Trichinella* muscle larvae in frozen wolf tissue in Italy. *J Parasitol* 75:472–473
52. Proulx J-F, MacLean JD, Gyorkos TW, Leclair D, Richter AK, Serhir B, Forbes L, Gajadhar AA (2002) Novel prevention program for trichinellosis in Inuit communities. *Clin Infect Dis* 34:1508–1514
53. Margolis HS, Midaugh JP, Burgess RD (1979) Arctic trichinosis: two Alaskan outbreaks from walrus meat. *J Infect Dis* 139:102–105
54. Forbes LB, Measures L, Gajadhar A, Kapel C (2003) Infectivity of *Trichinella nativa* in traditional northern (country) foods prepared with meat from experimentally infected seals. *J Food Prot* 66:1857–1863
55. Moravec F (2001) Redescription and systematic status of *Capillaria philippinensis*, an intestinal parasite of human beings. *J Parasitol* 87:161–164
56. Lu L-H, Lin M-R, Choi W-M, Hwang K-P, Hsu Y-S, Bair M-J, Liu J-D, Want T-E, Liu T-P, Chung W-C (2006) Human intestinal capillariasis (*Capillaria philippinensis*) in Taiwan. *Am J Trop Med Hyg* 74:810–813
57. Saichua P, Nithikathkul C, Kaewpitoon N (2008) Human intestinal capillariasis in Thailand. *World J Gastroenterol* 14:506–510
58. Cross JH (1992) Intestinal capillariasis. *Clin Microbiol Rev* 5:120–129
59. Beaver PC, Theis JH (1979) Dioctophymatid larval nematode in a subcutaneous module from man in California. *Am J Trop Med Hyg* 28:206–212
60. Le Bailly M, Leuzinger U, Bouchet F (2003) Dioctophymidae eggs in coprolites from Neolithic site of Arbon-Bleiche 3 (Switzerland). *J Parasitol* 89:1073–1076
61. Moore JG, Fry GF, Englert E Jr (1969) Thorny-headed worm infection in North American prehistoric man. *Science* 163:1324–1325
62. Schmidt GD (1971) Acanthocephalan infections of man, with two new records. *J Parasitol* 57:582–584
63. Tada I, Otsuji Y, Kamiya H, Mimori T, Sakaguchi Y, Makizumi S (1983) The first case of a human infected with an acanthocephalan parasite, *Bolbosoma* sp. *J Parasitol* 69: 205–208
64. Nuorteva P (1966) *Corynosoma strumosum* (Rudolphi) and *C. semerne* (Forssell) (Acanthocephala) as pathogenic parasites of farmed minks in Finland. *J Helminthol* 40:77–80
65. Buckner RL, Overstreet RM, Heard RW (1978) Intermediate hosts for *Tegorhynchus furcatus* and *Dollfusentis chandleri* (Acanthocephala). *Proc Helminthol Soc Wash* 45:195–201
66. Hitchcock DJ (1950) Parasitological study on the Eskimos in the Bethel area of Alaska. *J Parasitol* 36:232–234
67. Berland B (1961) Nematodes from some Norwegian marine fishes. *Sarsia* 2:1–50
68. Davey JT (1971) A revision of the genus *Anisakis* Dujardin, 1845 (Nematodes: Ascaridata). *J Helminthol* 45:51–72
69. Mattiucci S, Nascetti G, Cianchi R, Paggi L, Arduino P, Margolis L, Brattley J, Webb SC, D'Amelio S, Orecchia P, Bullini L (1997) Genetic and ecological data on the *Anisakis simplex* complex with evidence for a new species (Nematoda, Ascaridoidea, Anisakidae). *J Parasitol* 83:401–416
70. Mattiucci S, Nascetti G (2008) Advances and trends in the molecular systematics of anisakid nematodes, with implications for their evolutionary ecology and host-parasite co-evolutionary processes. *Adv Parasitol* 66:47–148

71. Yoshimura H, Akao N, Kondo K, Ohnishi Y (1979) Clinicopathological studies on larval anisakiasis, with special reference to the report of extra-gastrointestinal anisakiasis. *Jap J Parasitol* 28:347–354
72. K  ie M, Berland B, Burt MDB (1995) Development to third-stage larvae occurs in the eggs of *Anisakis simplex* and *Pseudoterranova decipiens* (Nematoda, Ascaridoidea, Anisakidae). *Can J Fish Aquat Sci* 52:134–139
73. Yoshinaga T, Kinami R, Hall KA, Ogawa K (2006) A preliminary study on the infection of anisakid larvae in juvenile greater amberjack *Seriola dumerilli* imported from China to Japan as Mariculture seedlings. *Fish Pathol* 41:123–126
74. Sakanari JA, McKerrow JH (1989) Anisakiasis. *Clin Microbiol Rev* 2:278–284
75. Audicana MT, Kennedy MW (2008) *Anisakis simplex*: from obscure infectious worm to inducer of immune hypersensitivity. *Clin Microbiol Rev* 21:360–379
76. Puente P, Anad  n AM, Rodero M, Romar  s F, Ubeira FM, Cu  llar C (2008) *Anisakis simplex*: the high prevalence in Madrid (Spain) and its relation with fish consumption. *Exp Parasitol* 118:271–174
77. Torres P, Jercic MI, Weitz JC, Dobrew EK, Mercado RA (2007) Human pseudoterranovosis, an emerging infection in Chile. *J Parasitol* 93:440–443
78. Bowen WD (1990) Population biology of sealworm (*Pseudoterranova decipiens*) in relation to its intermediate and seal hosts. *Canadian Bulletin of Fisheries and Aquatic Sciences* 222. Canadian Government Publishing, Canada
79. Jackson CJ, Marcogliese DJ, Burt MDB (1997) Role of hyperbenthic crustaceans in the transmission of marine helminth parasites. *Can J Fish Aquat Sci* 54:815–820
80. Ishikura H, Takahashi S, Sato N, Matsuura A, Nitto H, Tsunokawa M, Kikuchi K (1996) Epidemiology of anisakidiosis and related human diseases and studies on parasites infecting marine mammals, fishes and squids. *Bull Mar Biomed Inst Sapporo Med Univ* 3:23–37
81. Fagerholm H-P, Overstreet RM (2008) Ascaridoid nematodes: *Contracaecum*, *Porrocaecum*, and *Baylisascaris*. In: Atkinson CT, Thomas JN, Hunter DB (eds) *Parasitic diseases of wild birds*. Wiley-Blackwell, Ames, pp 413–433
82. Im KI, Shin HJ, Yoag TS (1989) Twenty cases of gastric anisakiasis. *Korean J Parasitol* 27:323 (Abstract, in Korean)
83. Vidal-Martinez VM, Osorio-Sarabia D, Overstreet RM (1994) Experimental infection of *Contracaecum multipapillatum* (Nematoda: Anisakinae) from Mexico in the domestic cat. *J Parasitol* 80:576–579
84. Petter AJ (1969) Enqu  te sur les n  matodes des sardines p  ch  es dans la r  gion nantaise. Rapport possible avec les granulomes   osinophiles observ  s chez l'homme dans la r  gion. *Ann Parasitol Hum Comp* 44:25–35
85. Petter AJ (1969) Enqu  te sur les nematodes des sardines p  ch  es dans la r  gion nantaise. Identification des larves d'ascarides parasitant les sardines (en rapport avec les granulomes   osinophiles observes chez l'homme dans la r  gion). *Ann Parasitol Hum Comp* 44:559–579
86. Overstreet RM, Meyer GW (1981) Hemorrhagic lesions in stomach of rhesus monkey caused by a piscine ascaridoid nematode. *J Parasitol* 67:226–235
87. Deardorff TL, Overstreet RM (1981) Larval *Hysterothylacium* (= *Thynnascaris*) (Nematoda: Anisakidae) from fishes and invertebrates in the Gulf of Mexico. *Proc Helminthol Soc Wash* 48:113–126
88. Gonzales L (1998) Experimental infection of mice with *Hysterothylacium aduncum* (Nematoda: Anisakidae) larvae from marine-farmed trout in Chile. *Arch Med Vet* 30:139–142
89. Yagi K, Nagasawa K, Ishikura H, Nakagawa A, Sato N, Kikuchi K, Ishikura H (1996) Female worm *Hysterothylacium aduncum* excreted from human: a case report. *Jpn J Parasitol* 45:12–23
90. Deardorff TL, Kliks MM, Desowitz RS (1983) Histopathology induced by larval *Terranova* (Type HA) (Nematoda: Anisakinae) in experimentally infected rats. *J Parasitol* 69:191–195
91. Nawa Y (1991) Historical review and current status of gnathostomiasis in Asia. *Southeast Asian J Trop Med Public Health* 22(Suppl):217–219
92. Diaz Camacho SP, Zazueta-Ramos M, Ponce-Torrecillas E, Osuna Ramirez I, Castro Velazquez R, Flores Gaxiola A, Baquera Heredia J, Willms K, Akahane H, Ogata K, Nawa Y (1998) Clinical manifestations and immunodiagnosis of gnathostomiasis in Culiacan, Mexico. *Am J Trop Med Hyg* 59:908–915
93. Schmutzhard E, Boongird P, Vejajiva A (1988) Eosinophilic meningitis and radiculomyelitis in Thailand, cause by CNS invasion of *Gnathostoma spinigerum* and *Angiostrongylus cantonensis*. *J Neurol Neurosurg Psychiatry* 51:80–87
94. Elzi L, Decker M, Battegay M, Rutishauser J, Blum J (2004) Chest pain after travel to the tropics. *Lancet* 363:1198
95. Ko R (1976) Experimental infection of mammals with larval *Echinocephalus sinensis* (Nematoda: Gnathostomatidae) from oysters (*Crassostrea gigas*). *Can J Zool* 54:597–609
96. Ko R (1977) Effects of temperature acclimation on infection of *Echinocephalus sinensis* (Nematoda: Gnathostomatidae) from oysters to kittens. *Can J Zool* 55:1129–1132
97. Goto Y, Tamura A, Ishikawa O, Miyachi Y, Ishii T, Akao N (1998) Creeping eruption caused by a larva of the suborder Spirurina type x. *Br J Dermatol* 139:315–318
98. Overstreet RM (2005) Medical importance: Infection by the rat lungworm, *Angiostrongylus cantonensis*. In: Rohde K (ed) *Marine parasitology*. CSIRO Publishing, Collingwood, Victoria, Australia, pp 442–446, 556–557
99. Overstreet R (2003) Flavor buds and other delights. *J Parasitol* 89:1093–1107
100. Measures LN (1988) Revision of the genus *Eustrongylides* J  gerski  ld, 1909 (Nematoda: Dioctophymatoidea) of piscivorous birds. *Can J Zool* 66:885–895
101. Mitchell AJ, Overstreet RM, Goodwin AE (2009) *Eustrongylides ignotus* infecting commercial bass, *Morone chrysops* female X *Morone saxatilis* male, and other fish in the southeastern USA. *J Fish Dis* 32:795–799
102. Beaver PC, Otsuji T, Otsuji A, Yoshimura H, Uchikawa R, Sato A (1983) Acanthocephalan, probably *Bolbosoma*, from

- the peritoneal cavity of man in Japan. *Am J Trop Med Hyg* 32:1016–1018
103. Choi C-J, Lee H-J, Go J-H, Park Y-K, Chia J-Y, Seo M (2010) Extraintestinal migration of *Centrorhynchus* sp. (Acanthocephala: Centrorhynchidae) in experimentally infected rats. *Korean J Parasitol* 48:139–143
 104. Schmidt GD (1972) Acanthocephala of captive primates. In: Fiennes TW, Karker S (eds) *Pathology of simian primates part II*. Karger, Basel, pp 144–156
 105. Dingley D, Beaver PC (1985) *Macracanthorhynchus ingens* from a child in Texas. *Am J Trop Med Hyg* 34:918–920
 106. Overstreet RM, Self JT, Vliet KA (1985) The pentastomid *Sebekia mississippiensis* sp. n. in the American alligator and other hosts. *Proc Helminthol Soc Wash* 52:266–277
 107. Mairena H, Solano M, Venegas W (1989) Human dermatitis caused by a nymph of *Sebekia*. *Am J Trop Med Hyg* 41:552–564
 108. Abadi MA, Stephney G, Factor SM (1996) Cardiac pentastomiasis and tuberculosis: the worm-eaten heart. *Cardiovasc Pathol* 5:169–174
 109. Dubey JP, Lunney JK, Shen SK, Kwok OCH, Ashford DA, Thulliez P (1996) Infectivity of low numbers of *Toxoplasma gondii* oocysts to pigs. *J Parasitol* 82:438–443
 110. Lindsay DS, Collins MV, Mitchell SM, Cole RA, Flick GJ, Wetch CN, Lindquist A, Dubey JP (2003) Sporulation and survival of *Toxoplasma gondii* oocysts in sea water. *J Eukaryot Microbiol* 50:5687–5688
 111. Fayer R, Dubey J, Lindsay D (2004) Zoonotic protozoa: from land to sea. *Trends Parasitol* 20:531–536
 112. Robertson LJ (2007) The potential for marine bivalve shellfish to act as transmission vehicles for outbreaks of protozoan infections in humans: a review. *Int J Food Microbiol* 120:201–216
 113. Lindsay DS, Phelps KK, Smith SA, Flick G, Summer SS, Dubey JP (2001) Removal of *Toxoplasma gondii* oocysts from sea water by eastern oysters (*Crassostrea virginica*). *J Eukaryot Microbiol* 48:S197–S198
 114. Arkush KD, Miller MA, Leutenegger CM, Gardner IA, Packham AE, Heckeroth AR, Tenter AM, Barr BC, Contad PA (2003) Molecular and bioassay-based detection of *Toxoplasma gondii* oocyst uptake by mussels (*Mytilus galloprovincialis*). *Int J Parasitol* 33:1087–1097
 115. Dubey JP, Zarnke R, Thomas JN, Wong SK, Van Bonn W, Briggs M, Davis JW, Ewing R, Mense M, Kwok OCH, Romand S, Thulliez P (2003) *Toxoplasma gondii*, *Neospora caninum*, *Sarcocystis neurona*, and *Sarcocystis canis*-like infection in marine mammals. *Vet Parasitol* 116:275–296
 116. Miller M, Conrad P, Gardner I, Kreuder C, Mazet J, Jessup D, Dodd E, Harris M, Ames J, Worcester K, Paradies D, Grigg M (2004) An update on *Toxoplasma gondii* infections in California sea otters. *Vet Parasitol* 125:133–134
 117. McDonald JC, Gyorkos TW, Alberton B, Maclean JD, Richer G, Juranek D (1990) An outbreak of toxoplasmosis in pregnant women in northern Quebec. *J Infect Dis* 161: 769–774
 118. Cali A, Takvorian PM (2003) Ultrastructure and development of *Pleistophora ronneae* n. sp., a Microsporidium (Protista) in the skeletal muscle of an immune-compromised individual. *J Eukaryot Microbiol* 50:77–85
 119. Fayer R (2004) Infectivity of microsporidia spores stored in seawater at environmental temperatures. *J Parasitol* 90:654–657
 120. Słodkiewicz-Kowalska A, Graczyk TK, Tamang L, Jedrzejewski S, Nowosad A, Zduniak P, Solarczyk P, Girouard AS, Majewska AC (2006) Microsporidian species known to infect humans are present in aquatic birds: implications for transmission via water? *Appl Environ Microbiol* 72:4540–4544
 121. Santín M, Fayer R (2009) *Enterocytozoon bienersi* genotype nomenclature based on the internal transcribed spacer sequence: a consensus. *J Eukaryot Microbiol* 56:34–38
 122. Rhinehart HL, Townsend FI, Overstreet RM, Visvesvara GS, da Silva A, Pieniazek NJ (1996) First report of microsporidiosis in the bottlenose dolphin, *Tursiops truncatus*. *Annual Conference of International Association for Aquatic Animal Medicine* 27:13–14
 123. Canning EU, Okamura B (2004) Biodiversity and evolution of the Myxozoa. *Adv Parasitol* 56:43–131
 124. Moncada L, López M, Murcia M, Nicholls S, León F, Guío O, Corredor A (2001) *Myxobolus* sp., another opportunistic parasite in immunosuppressed patients? *J Clin Microbiol* 39: 1938–1940
 125. Boreham R, Hendrick S, O'Donoghue P, Stenzel D (1998) Incidental finding of *Myxobolus* spores (Protozoa: Myxozoa) in stool samples from patients with gastrointestinal symptoms. *J Clin Microbiol* 36:3728–3730
 126. McClelland R, Murphy D, Cone D (1997) Report of spores of *Henneguya salminicola* (Myxozoa) in human stool specimens: possible source of confusion with human spermatozoa. *J Clin Microbiol* 35:2815–2818
 127. Lebbad M, Wilcox M (1998) Spores of *Henneguya salminicola* in human stool specimens. *J Clin Microbiol* 36:1820
 128. Venable DL, Gaudé AP III, Klerks PL (2000) Control of the trematode *Bolbophorus confusus* in channel catfish *Ictalurus punctatus* ponds using salinity manipulation and polyculture with black carp *Mylopharyngodon piceus*. *J World Aquacult Soc* 31:158–166
 129. Walker J (2005) Medical importance: marine schistosome dermatitis. In: Rhode K (ed) *Marine parasitology*. CABI, Wallingford, pp 439–442, 555–556
 130. Grodhaus G, Keh B (1958) The Marine dermatitis-producing cercaria of *Austrotilharzia variglandis* in California (Trematoda: Schistosomatidae). *J Parasitol* 44:633–638
 131. Chu GWTC, Cutress CE (1954) *Austrotilharzia variglandis* (Miller and Northup, 1926) Penner, 1953, (Trematoda: Schistosomatidae) in Hawaii with notes on its biology. *J Parasitol* 40:515–552
 132. Stunkard HW, Hinchliffe MC (1952) The morphology and life-history of *Microbilharzia variglandis* (Miller and Northup, 1926) Stunkard and Hinchliffe, 1951, avian blood-flukes whose larvae cause "swimmer's itch" of ocean beaches. *J Parasitol* 38:248–265

133. Barber KE, Caira JN (1995) Investigation of the life cycle and adult morphology of the avian blood fluke *Austrobilharzia variglandis* (Trematoda: Schistosomatidae) from Connecticut. *J Parasitol* 81:584–592
134. Bearup AJ (1955) A schistosome larva from the marine snail *Pyrazus australis* as a cause of cercarial dermatitis in man. *Med J Aust* 1:955–958
135. Overstreet RM, Curran SS (2005) Parasites of the American white pelican. *Gulf Caribbean Res* 17:31–48
136. Brant SV, Cohen AN, James D, Hui L, Hom A, Loker ES (2010) Cercarial dermatitis transmitted by exotic marine snail. *Emerg Infect Dis* 16:1357–1365
137. Leigh WH (1953) *Cercaria huttoni*, sp. nov., a dermatitis-producing schistosome larva from the marine snail, *Haminoea antillarum guadalupensis* Sowerby. *J Parasitol* 39:625–629
138. Leigh WH (1955) The Morphology of *Gigantobilharzia huttoni* (Leigh, 1953) an avian schistosome with marine dermatitis-producing larvae. *J Parasitol* 41:262–269
139. Nobile L, Fioravanti ML, Pampiglione S, Calderan M, Marchese G (1996) Report of *Gigantobilharzia acotylea* (Digenea: Schistosomatidae) in silver gulls (*Larus argentatus*) of the Venice Lagoon: considerations on its possible etiological role in the human dermatitis observed in the same area. *Parassitologia* 38:267
140. Penner L (1953) The biology of a marine dermatitis producing schistosome cercaria from *Batillaria minima*. *J Parasitol* 39:19–20
141. Malek EA, Chang TC (1974) Medical and economic malacology. Academic, New York
142. Penner L (1950) *Cercaria littoralinae* sp. nov., a dermatitis-producing schistosome larva from the marine snail, *Littorina planaxis* Philippi. *J Parasitol* 36:466–472
143. Ewers WH (1961) A new intermediate host of schistosome trematodes from New South Wales. *Nature* 190:283–284
144. Komiya Y, Ito J (1952) The morphology of *Cercaria sturniae* Tanabe, 1948 (cercaria of *Gigantobilharzia sturniae* Tanabe, 1951), a cause of cercaria dermatitis in Japan. *Jpn J Med Sci Biol* 5:215–220
145. Kolárová L (2007) Schistosomes causing cercarial dermatitis: a mini-review of current trends in systematics and of host specificity and pathogenicity. *Folia Parasitol* 54:81–87
146. Hrádková K, Horák P (2002) Neurotopic behaviour of *Trichobilharzia regenti* in ducks and mice. *J Helminthol* 76:137–141
147. Kolárová L, Horák P, Cada F (2001) Histopathology of the CNS and nasal infections caused by *Trichobilharzia regenti* in vertebrates. *Parasitol Res* 87:644–650
148. Little MD (1965) Dermatitis in a human volunteer infected with *Strongyloides* of nutria and raccoon. *Am J Trop Med Hyg* 14:1007–1009
149. Deardorff TL, Overstreet RM, Okihiro M, Tam R (1986) Piscine adult nematode invading an open lesion in a human hand. *Am J Trop Med Hyg* 35:827–830
150. Recavarren-Arce S, Velarde C, Gotuzzo E, Cabrera J (1999) Amoeba angeitic lesions of the central nervous system in *Balamuthia mandrillaris* amoebiasis. *Hum Pathol* 30:269–273
151. Sawyer TK, Visvesvara GS, Harke BA (1977) Pathogenic amoebas from brackish water and ocean sediments with a description of *Acanthamoeba hatchetti*, n. sp. *Science* 196:1324–1325
152. Daggett P-M (1982) Protozoa from polluted waters; potential human pathogens. In: Colwell RR (ed) *Microbial hazards of diving in polluted waters: a proceedings*. Maryland Sea Grant Publication UM-SG-TS-82-01. University of Maryland, College Park, pp 39–42
153. Lorenzo-Morales J, Monteverde-Miranda CA, Jiménez C, Tejedor ML, Valladares B, Ortega-Rivas A (2005) Evaluation of *Acanthamoeba* isolates from environmental sources in Tenerife, Canary Islands, Spain. *Ann Agric Environ Med* 12:233–236
154. Fernandez M-CA, Crespo EP, Mallen MM, Ares MPMP, Casas MC (1989) Marine amoebae from waters of northwest Spain, with comments on a potentially pathogenic euryhaline species. *J Eukaryot Microbiol* 36:239–241
155. Ledee DR, Hay J, Byers TJ, Seal DV, Kirkness CM (1996) *Acanthamoeba griffini*: molecular characterization of a new corneal pathogen. *Invest Ophthalmol Vis Sci* 37:544–550
156. Lorenzo-Morales J, Coronado-Álvarez N, Martínez-Carretero E, Maciver SK, Valladares B (2007) Detection of four adenovirus serotypes within water-isolated strains of *Acanthamoeba* in the Canary Islands, Spain. *Am J Trop Hyg* 77:753–756
157. Mattana A, Serra C, Mariotti E, Delogu G, Fiori PL, Cappuccinelli P (2006) *Acanthamoeba castellanii* promotion of in vitro survival and transmission of coxsackie B3 viruses. *Eukaryot Cell* 5:665–671
158. Laskowske-Arce MA, Orth K (2008) *Acanthamoeba castellanii* promotes the survival of *Vibrio parahaemolyticus*. *Appl Environ Microbiol* 74:7183–7188
159. Overstreet RM, Jovonovich J, Ma H (2009) Parasitic crustaceans as vectors of viruses, with an emphasis on three penaeid viruses. *Integr Comp Biol* 49:127–141
160. Nuzzi R, Zaki MH (1982) Unusual health effects associated with surface waters. *New York State J Med* 82:1347–1349
161. Johnson D, Reynolds K, Gerba C, Pepper I, Rose J (1995) Detection of *Giardia* and *Cryptosporidium* in marine waters. *Water Sci Technol* 31:439–442
162. Thompson R (2004) The zoonotic significance and molecular epidemiology of *Giardia* and giardiasis. *Vet Parasitol* 126:15–35
163. Monis P, Andrews R, Mayrhofer G, Ey P (2003) Genetic diversity within the morphological species *Giardia intestinalis* and its relationship to host origin. *Infect Genet Evol* 3:29–38
164. Gaydos J, Miller W, Gilardi K, Melli A, Schwantje H, Engelstoft C, Fritz H, Conrad P (2007) *Cryptosporidium* and *Giardia* in marine-foraging river otters (*Lontra canadensis*) from the Puget Sound Georgia Basin ecosystem. *J Parasitol* 93:198–202

165. Olson M, Appelbee A, Measures L (2004) *Giardia duodenalis* and *Cryptosporidium parvum* infections in pinnipeds. *Vet Parasitol* 125:131–132
166. Gaydos J, Miller W, Jognson C, Zornetzer H, Melli A, Packham A, Jeffries S, Lance M, Conrad P (2008) Novel and canine genotypes of *Giardia duodenalis* in Harbor Seals (*Phoca vitulina richardsi*). *J Parasitol* 94:1264–1268
167. Johnson D, Enriquez C, Pepper I, Davis T, Gerba C, Rose J (1997) Survival of *Giardia*, *Cryptosporidium*, Poliovirus and *Salmonella* in marine waters. *Water Sci Technol* 35:261–268
168. Schets F, Van den Berg H, Engels G, Lodder W, de Roda HA (2007) *Cryptosporidium* and *Giardia* in commercial and non-commercial oysters (*Crassostrea gigas*) and water from the Oosterschelde, the Netherlands. *Int J Food Microbiol* 113:189–194
169. Lucy FE, Graczyk TK, Tamang L, Miraflor A, Minchin D (2008) Biomonitoring of surface and coastal water for *Cryptosporidium*, *Giardia*, and human-virulent microsporidia using molluscan shellfish. *Parasitol Res* 103:1369–1375
170. Fayer R, Graczyk T, Lewis E, Trout J, Farley C (1998) Survival of infectious *Cryptosporidium parvum* oocysts in seawater and eastern oysters (*Crassostrea virginica*) in the Chesapeake Bay. *Appl Environ Microbiol* 64:1070–1074
171. Bogomolni AL, Gast RJ, Ellis JC, Dennett M, Pugliares KR, Lentell BJ, Moore MJ (2008) Victims or vectors: a survey of marine vertebrate zoonoses from coastal waters of the Northwest Atlantic. *Dis Aquat Org* 81:13–38
172. Fayer R (2004) *Cryptosporidium*: a water-borne zoonotic parasite. *Vet Parasitol* 126:37–56
173. Appelbee A, Thompson R, Olson M (2005) *Giardia* and *Cryptosporidium* in mammalian wildlife – current status and future needs. *Trends Parasitol* 21:370–376
174. Dupont HL, Chappell CL, Sterling CR, Okhuysen PC, Rose JB, Jakubowski W (1995) The infectivity of *Cryptosporidium parvum* in healthy volunteers. *N Engl J Med* 332:855–859
175. Graczyk T, Marcogliese D, de LaFontaine Y, Da Sliva A, Mhangami-Ruwende B, Pieniazek N (2001) *Cryptosporidium parvum* oocysts in zebra mussels (*Dreissena polymorpha*): evidence from the St Lawrence River. *Parasitol Res* 87:231–234
176. Hughes-Hanks J, Rickard L, Panuska C, Saucier J, O'Hara T, Dehn L, Rolland R (2005) Prevalence of *Cryptosporidium* spp. and *Giardia* spp. in five marine mammal species. *J Parasitol* 91:1225–1228
177. Santín M, Dixon B, Fayer R (2005) Genetic characterization of *Cryptosporidium* isolates from ringed seals (*Phoca hispida*) in Northern Quebec, Canada. *J Parasitol* 91:712–716
178. Graczyk TK, Sunderland D, Awantang GN, Mashinski Y, Lucy FE, Graczyk Z, Chomicz L, Breyse PN (2010) Relationships among bather density, levels of human waterborne pathogens, and fecal coliform counts in marine recreational beach water. *Parasitol Res* 106:1103–1108
179. Graczyk TK, Majewska AC, Schwab KJ (2008) The role of birds in dissemination of human waterborne enteropathogens. *Trends Parasitol* 24:55–59
180. Ryan U (2010) *Cryptosporidium* in birds, fish and amphibians. *Exp Parasitol* 124:113–120
181. Armentia A, Martín-Gil FJ, Pascual C, Martín-Esteban M, Callejo A, Martínez C (2006) *Anisakis simplex* allergy after eating chicken meat. *J Investig Allergol Clin Immunol* 16:258–263
182. Del Pozo MD, Moneo I, Fernández de Corres L, Audicana MT, Munoz D, Fernandez E, Navarro JA, Garcia M (1996) Laboratory determinations in *Anisakis simplex* allergy. *J Allergy Clin Immunol* 97:977–984
183. Del Pozo MD, Audicana MT, Diez J, Muñoz I, Ansotegui IJ, Fernández E, García M, Etxenaguisa M, Monio I, Fernández de Corres L (1997) *Anisakis simplex*, a relevant etiologic factor in acute urticaria. *Allergy* 52:576–576
184. Audicana MT, Ansotegui IJ, Fernández de Corres L, Kennedy MW (2002) *Anisakis simplex*: dangerous – dead and alive? *Trends Parasitol* 18:20–25
185. Daschner A, Alonso-Gómez A, Cabañas R, Suarez-d-Parga J-M, López-Serrano M-C (2000) Gastroallergic anisakiasis: borderline between food allergy and parasitic disease – Clinical and allergologic evaluation of 20 patients with confirmed acute parasitism by *Anisakis simplex*. *J Allergy Clin Immunol* 105:176–181
186. Alonso-Gómez A, Moreno-Ancillo A, López-Serrano MC, Suarez-de-Parga JM, Daschner A, Caballero MT, Barranco P, Cabañas R (2004) *Anisakis simplex* only provokes allergic symptoms when the worm parasitizes the gastrointestinal tract. *Parasitol Res* 93:378–384
187. Takahashi S, Ishikura H, Sato M, Iwasa K (1993) Seroimmunodiagnosics of anisakiasis. *Ann Rev Immunol Cyugai-Medical, Tokyo*, 211–217 (in Japanese)
188. Nieuwenhuizen N, Lopata AL, Jeebhay MF, Herbert DR, Robins TG, Brombacher F (2006) Exposure to the fish parasite *Anisakis* causes allergic airway hyperreactivity and dermatitis. *J Allergy Clin Immunol* 117:1098–1105
189. Jeebhay MF, Robins TG, Miller ME, Bateman E, Smuts M, Baatjies R, Lopata AL (2008) Occupational allergy and asthma among salt water fish processing workers. *Am J Ind Med* 51:899–910
190. Vázquez-López C, de Armas-Serra C, Bernardina W, Rodríguez-Cabeiro F (2001) Oral inoculation with *Gymnorhynchus gigas* induces anti-parasite anaphylactic [sic] antibody production in both mice and rats and adverse reactions in challenge mice. *Int J Food Microbiol* 64:307–315
191. U.S. Food and Drug Administration (1997) Food code, sections 3–401.11 (A1) and 3–401.12. In: 1997 Recommendations of the United States Public Health Administration. U.S. Food and Drug Administration, Washington, D.C.
192. Adams AM, Miller KS, Wekell MM, Dong FM (1999) Survival of *Anisakis simplex* in microwave-processed arrowtooth flounder (*Atheresthes stomias*). *J Food Prot* 62:403–409
193. EFSA Panel on Biological Hazards (BIOHAZ) (2010) Scientific opinion on risk assessment of parasites in fishery products. *EFSA Journal* 8:1543 (91 pp)

194. Levsen A, Lunestad BT, Berland B (2005) Low detection efficiency of candling as a commonly recommended inspection method for nematode larvae in the flesh of pelagic fish. *J Food Protect* 68:828–832
195. Karl H, Leinemann M (1993) A fast and quantitative detection method for nematodes in fish fillets and fishery products. *Arch Lebensmittelhyg* 44:124–125
196. Heia K, Sivertsen AH, Stormo SK, Wold JP, Nilsen H (2007) Detection of nematodes in cod (*Gadus morhua*) fillets by imaging spectroscopy. *J Food Sci* 72:E11–E15

Books and Reviews

- Anderson RC (2000) Nematode parasites of vertebrates: their development and transmission, 2nd edn. CABI, New York
- Beaver PC, Jung RC, Cupp EW (1984) Clinical parasitology, 9th edn. Lea and Febiger, Philadelphia
- Brusca RC, Brusca GJ (2002) Invertebrates, 2nd edn. Sinauer, Sunderland, MA
- Coombs I, Crompton DWT (1991) A guide to human helminths. Taylor & Francis, London
- Crompton DWT, Saviolo L (2007) Handbook of helminthiasis for public health. CRC Press, Boca Raton, FL
- Eiras JC, Segner H, Wahli T, Kapoor BG (eds) (2008) Fish diseases, volumes 1 and 2. Science Publishers, Enfield, NH
- Fayer R (2010) Taxonomy and species delimitation in *Cryptosporidium*. *Exp Parasitol* 124:90–97
- Garcia LS (2007) Diagnostic medical parasitology, 5th edn. ASM Press, Washington, DC
- Grabda J (1991) Marine fish parasitology: an outline. VCH, New York
- Horák P, Kolářová L, Adema CM (2002) Biology of the schistosome genus *Trichobilharzia*. *Adv Parasitol* 52:155–233
- Krauss H, Weber A, Appel M, Enders B, Isenbert HD, Schiefer HG, Slenczka W, von Graevenitz A, Zahner H (2003) Zoonoses: infections diseases transmissible from animals to humans, 3rd edn. ASM Press, Washington, DC
- Murrell KD, Fried B (2007) World class parasites: volume 11, Food-borne parasitic zoonoses, fish and plant-borne parasites. Springer, New York
- Nawa Y, Nakamura-Uchiyama F (2004) An overview of gnathostomiasis in the world. *Southeast Asian J Trop Med Public Health* 35(Suppl 1):87–91
- Orihel TC, Ash LR (1995) Parasites in human tissues. American Society of Clinical Pathologists, Chicago
- Rhode K (ed) (2005) Marine parasitology. CABI, New York
- Roberts LS, Janovy J (2008) Foundations of parasitology, 8th edn. McGraw-Hill, New York
- Williams H, Jones A (1994) Parasitic worms of fish. Taylor & Francis, London
- Woo PTK (ed) (2006) Fish diseases and disorders vol 1: Protozoan and metazoan infections, 2nd edn. CABI, Cambridge, MA
- Zhou X-N, Lv S, Yang G-J, Kristensen TK, Bergquist NR, Utzinger J, Maline JB (2009) Spatial epidemiology in zoonotic parasitic diseases: insights gained at the 1st International Symposium on Geospatial Health in Lijiang, China, 2007. *Parasit Vectors* 2:10–25

Weather Prediction Models

JULIO T. BACMEISTER

NCAR Earth System Laboratory, National Center for Atmospheric Research, Boulder, CO, USA

Article Outline

Glossary

Definition of the Problem

Introduction: Direct Simulation of Atmospheric Flows

Development of Modern NWP Models

How Are NWP Models (Versus Climate Models) Evaluated?

Initialization and Data Assimilation

Future Directions

Bibliography

Glossary

Assimilation The process of combining observations of the atmosphere with a “first guess” (usually a model forecast) to define the atmospheric state on a forecast model grid.

Geostrophic balance A possible state of rotating fluids in which flow is directed along pressure gradients rather than across them.

Gravity waves Rapidly moving atmospheric disturbances driven by gravity acting on vertical density gradients. Often arise as a consequence of spurious geostrophic imbalance in initial conditions.

Hydrostatic balance State in which the vertical pressure gradient force cancels the downward acceleration of gravity. Approximately obeyed in atmospheric flows with horizontal scales larger than several km.

Instabilities (or unstable modes) Spatial patterns in a flow that are able to extract energy from the background flow and grow in amplitude.

Primitive equations Complete set of equations describing flow of a thin envelope of fluid or gas surrounding a sphere.

Resolution Separation in space of notional points at which quantities are defined in a numerical model.

Definition of the Problem

Awareness of weather and concern about weather in the proximate future certainly must have accompanied the

emergence of human self-consciousness. Although weather is a basic idea in human existence, it is difficult to define precisely. Weather intuitively refers to a set of atmospheric conditions prevailing over a relatively small area, and even more emphatically, over a relatively short time. The immediacy contained in our notion of weather may be reflected in the fact that in many languages the same root appears in the words for time and weather, for example, Spanish (*tiempo*) or Hungarian (*idő*). Thus, weather is to be distinguished from the notion of climate, or more subtly from the notion of “spells” which imply a time window anywhere from a week to several years. Our experience of weather does involve quantities which can be defined with reasonable precision. These include air temperature, wind speed, precipitation rates and types, cloud cover, and also humidity, air-quality, and barometric pressure. Numerical weather prediction (NWP) is the attempt to predict the evolution of such quantities by solving a set of partial differential equations which describe the dynamics of a fluid like the atmosphere [1, 2]. These equations must be solved using approximate or “numerical” techniques using computers. They are integrated forward in time from a set of initial conditions, which are derived from an optimized combination of observations and previous model forecasts. This bootstrapped procedure is known as the analysis cycle and is an integral part of the activities at all modern forecasting centers.

NWP and global climate simulation are closely related problems. The models used in both endeavors are essentially the same. A key difference between NWP and the climate problem is the role of atmospheric initial conditions. Initialization of the atmosphere is of secondary importance in multiyear climate simulations. However, good initialization of the atmospheric state is at the heart of the forecasting problem. Initialization must give an accurate and comprehensive representation of the state of the atmosphere that is compatible with numerical forecast model being used. The initial state must also satisfy a number of “balance constraints” to avoid spurious initial variability.

Lewis Fry Richardson (1881–1953) reported the first numerical weather forecast, performed using hand calculations, in 1922 [3]. His attempt did not succeed for reasons that are summarized below, and discussed in detail in the book by Lynch [2], but

Richardson’s effort marks the beginning of NWP as a field of inquiry.

Introduction: Direct Simulation of Atmospheric Flows

In order to appreciate the challenges faced by numerical models of the global atmosphere it is useful to have a sense of the complex nature of the motions which must be represented. A brief account of some of the dynamical processes that play a part in weather is given below. For more details, the reader is referred to the excellent, comprehensive introduction to the dynamics of the atmosphere and ocean by A.E. Gill (1937–1986) [4].

A Zoology of Atmospheric Motion

The atmospheric flows which are responsible for creating weather span a wide range of space and time scales and are driven by rich variety of dynamical and thermodynamic processes. Atmospheric flows may be forced by features at the Earth’s surface. Mountainous terrain forces atmospheric circulations with spatial scales ranging from several kilometers to several thousand kilometers. Surface temperature contrasts, most pronounced between land and ocean but also created by variations in ground cover, force sea-breeze circulations with scales of tens of kilometers, [5] as well as continental-scale monsoonal circulations (Table 1) [6–8].

External forcing can produce wave-like motions in the atmosphere. Gravity waves, or buoyancy waves, exist because of density stratification in the atmosphere, and are analogs to waves on the surface of water. These waves are typically one to hundreds of kilometers in scale, and have periods of minutes to hours, although in the tropics both spatial and temporal scales may be longer [4]. Gravity waves may play a role in triggering convection [9–11], as well as in organizing convection in the mesoscale [10, 12, 13]. Mountain waves are large, nonlinear, gravity waves generated by flow over mountains with horizontal scales from less than ten kilometers to several hundred kilometers. Such waves are responsible for a number of local but intense winds [14]. Gravity waves are also notorious for contaminating forecasts when errors in initial conditions are present.

Weather Prediction Models. Table 1 Motions in atmosphere

Phenomenon	Spatial scale	Temporal scale	Role in weather
Acoustic waves	Meters	Seconds	None
Gravity waves	1–1,000 km	15 min to days	Initiating and organizing convection. Mountain flows, e.g., chinooks, Foehns . . .
Easterly waves	500–2,000 km	Days	Organizing tropical convection, tropical cyclogenesis
Baroclinic instability	Thousands of km	Days	Midlatitude cyclogenesis
Madden–Julian oscillation	10,000 km	20–60 days	Possible modulation of tropical and midlatitude weather

Rossby waves [4, 15–17] named after the Swedish meteorologist Carl Gustav Rossby exist because of the change in the effective rotation rate experienced by a fluid parcel as it moves from equator to pole. The effective rotation rate at the Equator is zero, since all of the effects of Earth’s rotation are directed in the vertical and are felt as a slight reduction in the pull of gravity. At the poles, the effective rotation rate is as intuitively expected – one cycle or 2π radians per day. In between equator and pole, this rate varies as the sine of latitude. Rossby waves which exploit this gradient have scales of thousands to ten thousands of kilometers and periods of days to weeks. Immense planetary-scale Rossby waves or “planetary waves” generated by topographic features such as the Himalayas and Rockies, as well as, by continental-scale land-sea contrasts, dominate the tropospheric flow and are responsible for the mean position of the jet-streams, and for the mean paths of storms.

Probably more important in the overall problem of weather, and certainly more difficult to predict, is another class of flows driven by internal exchanges of

energy in the atmosphere. These instabilities or unstable modes are essentially flow patterns that are able to extract energy from their surroundings and grow in amplitude. Initial growth of unstable modes is typically exponential. Familiar examples of such behavior in fluids include the growth of wind driven waves on the surface of water. The convective instability of a fluid heated from below is another example that is both familiar in everyday experience and also important in the atmosphere.

In the atmosphere, a particularly important mode of unstable growth is through baroclinic instability [18, 19]. This instability arises from a combination of thermal and inertial effects in a rotating fluid with a horizontal temperature along the lower boundary. Baroclinic instability is characterized by length scales of 1,000 km and growth times of days – making it a key factor in weather. It is fair to say that the problem of weather forecasting in midlatitudes is essentially that of predicting the evolution of baroclinically unstable modes in the atmosphere.

The tropics possess another as yet poorly understood class of motions, in which moist heating plays a key role in energizing and modifying wave motions in the atmosphere [20]. Tropical easterly waves [21] have periods of several days and scales of hundreds to thousands of kilometers and play an important role in the genesis of tropical cyclones [22–24]. The Madden–Julian oscillation or “MJO” [25, 26] is an eastward traveling disturbance in the tropics with a length scale close to 10,000 km and a period of weeks. It is thought to play a role in modulating tropical cyclone frequency in various basins [27–31] and possibly midlatitude disturbances as well [32, 33]. The dynamics behind the MJO are not yet understood. Successful forecasting of the MJO could improve prospects for accurate forecasts out to lead times of weeks [34].

Early History

Weather prediction, not climate simulation, was the original motivation for developing numerical models that describe the time evolution of the atmosphere. Richardson’s 1922 attempt at NWP predates initial attempts to study climate numerical models by at least four decades. The first attempts at simulating the longer-term equilibrium state of the atmosphere did not occur until the 1960s [35].

The notion of predicting weather systematically using equations to describe the evolution of systems can probably be traced to the Norwegian meteorologist Vilhelm Bjerknes (1862–1951) [1, 2] who founded the famous “Bergen School” of meteorology [36, 37]. Bjerknes considered graphical methods to predict the motion of fronts and other features in the atmosphere, as well as numerical techniques to solve the equations themselves. However, it was L. F. Richardson who finally conceived and implemented a concrete plan to use a numerical solution of the partial differential equations describing the atmosphere (see section on “primitive equations”) to make forecasts. His approach was remarkably prescient both in concept and in detail. He employed a finite-difference technique on a regularly spaced grid of points over central Europe and attempted to predict the tendency of surface pressure over 6 h. Richardson’s forecast was a famous failure or “bust” (see the book by Lynch [2] for a detailed and readable account of Richardson’s attempt, as well as for a comprehensive account of the development of NWP). However, the reasons for Richardson’s failure lay in the initial conditions used in the forecast not in his method as Richardson himself suspected [2].

The potential for numerical prediction was clear. The major obstacle beside the question of initialization, was the sheer amount of calculation required to produce even a short forecast over a limited area. Richardson imagined computational “factories” employing thousands of people to produce weather forecasts [2]. The appearance of electronic computers soon after World War II made numerical prediction plausible. The potential application to the problem of weather prediction was recognized by one of the main intellects behind the development of electronic computers John von Neumann (1903–1957) [1]. In the first successful attempts at NWP using electronic computers so-called filtered equation sets were used [2]. Filtered equations describe a limited set of atmospheric motions, but allow large time steps to be taken in numerical integration and side step the need for well-balanced initial conditions (see sections on “Numerics and Initialization”).

Development of Modern NWP Models

It was recognized early on by Jule Charney and others that models using filtered equations were not a

promising long-term path for NWP [2]. As computer power increased, the limitations on the time-step length allowed by more complete equations became less important. The problem of initialization was not solved, but its tractability became apparent [38]. Development work on NWP models using the primitive equations began in the late 1950s and eventually led to the adoption of primitive equation models at all major forecasting centers by the mid 1960s [2].

The Primitive Equations

The primitive equations are essentially the complete Navier–Stokes or Euler equations for fluid motion with the hydrostatic approximation invoked. As long as the horizontal scales of interest are much larger than 10 km, the hydrostatic approximation is well satisfied. However, at the time of writing, major NWP models are reaching horizontal resolutions that test the limits of this approximation. The Euler equations using a generalized vertical coordinate η , including connections between the fully nonhydrostatic system and the hydrostatic system currently used in NWP models, are nicely described by Laprise [39].

The hydrostatic primitive equations include an equation describing the evolution of horizontal momentum or velocity V :

$$\frac{d}{dt} V + f \mathbf{k} \times V = -\alpha \nabla_{\eta} p - \nabla_{\eta} \phi + F_{phys}$$

where f is the Coriolis parameter or local apparent rotation rate and $\mathbf{k}$ is the unit vector in vertical direction. The symbol ∇_{η} denotes a gradient along surfaces of fixed η . As is common in the meteorological literature ϕ denotes the geopotential height or potential energy density of a fluid parcel along constant η . The remaining symbols α and p denote specific volume and pressure. This equation is simply the fluid dynamical form of Newton’s law $F = ma$, where the right-hand side contains forces accelerating fluid parcels in the horizontal. In current meteorological literature, the individual velocity components are usually designated as u for the eastward or “zonal” component, and v for the northward or “meridional” component.

Another equation restates the first law of thermodynamics $dE = dQ + dW$ in fluid form:

$$C_p \frac{d}{dt} T - \alpha \frac{d}{dt} p = H_{phys}$$

where T is the absolute temperature and C_p is the heat capacity of air at constant pressure. In both the momentum and energy equations the symbol $\frac{d}{dt}$ is used to denote:

$$\left(\frac{\partial}{\partial t}\right)_\eta + V \cdot \nabla_\eta + \dot{\eta} \frac{\partial}{\partial \eta}$$

the Lagrangian derivative that tracks changes in a quantity following a fluid parcel.

A prognostic equation for mass continuity is also required:

$$\left[\frac{\partial}{\partial t} \left(\frac{\partial p}{\partial \eta}\right)\right]_\eta + \nabla_\eta \cdot \left(V \frac{\partial p}{\partial \eta}\right) + \frac{\partial}{\partial \eta} \left(\dot{\eta} \frac{\partial p}{\partial \eta}\right) = 0$$

where $g^{-1} \frac{\partial p}{\partial \eta}$ is the mass per unit area in a column between surfaces of constant η .

Three diagnostic relationships are needed to complete the system. The equation of state for a gas:

$$p\alpha = RT$$

and a relationship that determines the geopotential height of η -surfaces:

$$\phi = \phi_s + \int_\eta^{\eta_s} \alpha \frac{\partial p}{\partial \eta'} d\eta'$$

This relationship uses the surface geopotential height ϕ_s as well as the integral of the hydrostatic relationship:

$$\frac{\partial p}{\partial \eta} = \frac{1}{\alpha} \frac{\partial \phi}{\partial \eta}$$

between pressure and specific volume.

The right-hand sides of the momentum and energy equations also contain the terms F_{phys} and H_{phys} . These represent the effects of physical parameterizations on the grid scale variables, and include effects from radiative heating, friction, and other processes which will be described in more detail in the section on [parameterization](#). The use of the generalized vertical coordinate η gives the equations a somewhat unfamiliar look. However, replacing η with the geometric height z , and noting that $\nabla_{\perp \rho} = 0$ reduces them to a more familiar form.

Due to the complex shape of Earth's topography most NWP models do not use geometric height as their

vertical coordinate. Most use a version of the so-called σ -coordinate defined by:

$$\sigma = \frac{p - p_t}{p_s(x, y, t) - p_t}$$

where $p_s(x, y, t)$ is the surface pressure and p_t is the pressure at the model top, typically a constant value. The coordinate surface $\sigma = 1$ follows the bottom of the model domain while $\sigma = 0$ follows the top. Boundary conditions on “vertical” velocity become simply $\dot{\sigma}(1) = \dot{\sigma}(0) = 0$. Thus, the difficulties of representing flow boundaries in and around topographic obstacles are replaced by the need for a prognostic equation describing π_s . This is obtained by integrating the mass continuity equation in the vertical.

The primitive equations can describe all of the motions discussed in the introduction except for fully three dimensional acoustic waves. They do allow a horizontal acoustic mode known as the Lamb wave [4] which can produce difficulties for numerical integrations. Gravity waves and convective instabilities with horizontal scales much smaller than 100 km are not well represented, and this may become a significant handicap for models in the next 10 years (see section on [“Future Directions”](#)).

In addition to the equations describing the dynamics and thermodynamics of the atmosphere, modern NWP models include equations that describe the evolution of trace gases and trace species in the atmosphere. Each additional species results in an additional prognostic equation of the form:

$$\frac{d}{dt} q_i = C_{i,phys}$$

where q_i is the mixing ratio of the i th species and $C_{i,phys}$ are the sources and losses of the species. The most important of these trace quantities is water vapor. Water vapor was included in primitive equation NWP models early on [40]. As NWP model domains were extended into the tropics during the 1970s [41] strong condensational heating associated with high tropical humidities presented problems for NWP that spurred the development of deep convection parameterizations [42] (see section on [“Parameterization”](#)). More recently, NWP models have incorporated prognostic treatment of condensed water species known as prognostic cloud schemes [43, 44].

Numerics

Closed form solutions of the primitive equations do not exist. Approximate numerical techniques must be used. An illustration of how this proceeds is given here using a simple equation that describes one dimensional advection of a constituent C by a constant flow u :

$$\partial_t C + u \partial_x C = S$$

Figure 1 shows three time steps from a numerical integration of this equation. Finite-difference approximations of the partial derivatives in both space and time are calculated as shown in the figure. The approach illustrated is known as a “centered difference” since the approximation uses a symmetric stencil of equally weighted points. With these approximations to the derivatives in time and space, a solution for the tracer distribution at $\tau + 1$ can be obtained:

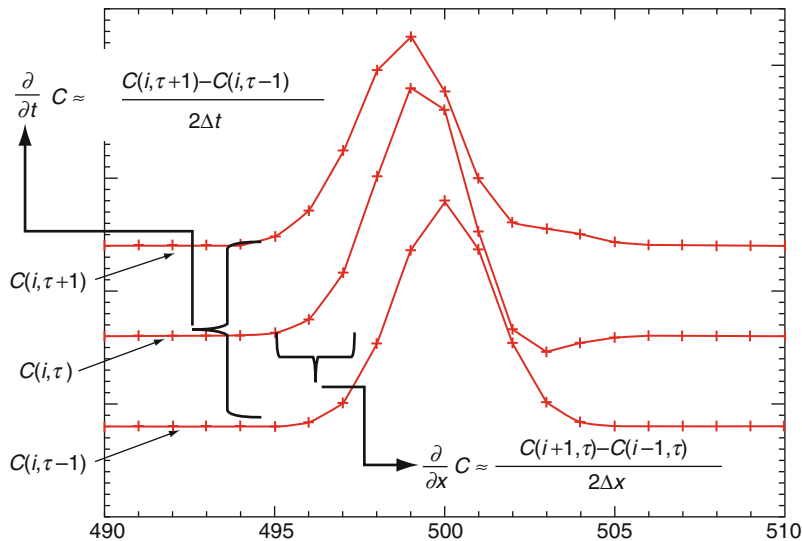
$$C(i, \tau + 1) \approx C(i, \tau - 1) - u \Delta t \left(\frac{C(i + 1, \tau) - C(i - 1, \tau)}{2 \Delta x} \right)$$

As a technical detail, notice that to start (or initialize) this calculation two time levels of data must be given. In practice, these can be set equal to each other.

Once initial values are given, the arithmetic equation above can be repeated, or iterated, many times to give the time evolution of C .

The algorithm described above is a simple but stable scheme that was employed in early numerical models. However, it has many undesirable properties, which are described in [45]. Nevertheless, the basic concepts illustrated by this method hold for all explicit finite-difference approaches. In implicit methods, the terms in the approximation of $\partial_x C$ above are replaced by their values at time $\tau + 1$. In the case above, this leads to a tridiagonal matrix problem that can be easily solved using standard techniques [45].

The most natural coordinate system for a global atmospheric model is spherical, with latitude and longitude as the horizontal coordinates. Unfortunately, as either pole is approached, equally spaced longitude lines come arbitrarily close together. This presents problems for most finite-difference numerical schemes, which become inaccurate or unstable when information travels across multiple grid lengths in a single time step. For example, in the simple case described in Fig. 1 the time step must be chosen such that $\frac{u \Delta t}{\Delta x} < 1$ to avoid numerical instability. Versions of this limit, known as the Courant-Friedrichs-Levy (CFL) limit, exist for



Weather Prediction Models. Figure 1

Three time steps from numerical advection of a Gaussian tracer pulse using a second-order, space-centered, time-centered finite-difference scheme. The +’s indicate grid-point locations. A constant velocity of $u = -0.5$ is used with grid spacing $\Delta x = 1$ and time step $\Delta t = 1$

most explicit finite-difference schemes. In systems of equations that support propagating waves, as the primitive equations do, the relevant velocity in the CFL limit is typically the sum of the wave propagation speeds and the advective speed. Thus, the stability of explicit calculations is limited by the fastest wave modes in the system, which are often of little interest, for example, Lamb waves or deep gravity waves. Global models whose numerics are based on finite differences (“grid-point models”) address this issue by introducing polar filters which are designed to suppress small-scale motions in the polar regions of the model.

A revolutionary innovation occurred in the late 1960s with the introduction of spectral techniques for solving the primitive equations [46]. Spectral models decompose the atmosphere into a finite sequence of spherical harmonic functions, rather than onto a grid of points. For example, the wind u is approximated by:

$$u(\lambda, \theta, z, t) = \sum_{n=0}^N \sum_{m=-n}^n U_n^m(z, t) Y_n^m(\lambda, \theta)$$

where λ is longitude and θ is latitude. The time evolution of the atmospheric flow is then represented as the time evolution of the amplitudes $U_n^m(z, t)$. Nonlinear advection terms arising in the primitive equations may be dealt with directly in spectral space, as interactions of spherical harmonics or in physical space with transformation of the tendencies back to spectral space [45].

The resolution of spectral models is determined by the truncation parameter N . The form of the finite sum used above is referred to as a triangular truncation. A model with $N = 799$ in this kind of decomposition is referred to as a “T799” model. Lynch [2] gives a useful rule-of-thumb for estimating the spatial resolution Δ corresponding a particular value of N . He uses $\Delta \approx (2\pi a_e)/2 N$ or $\Delta \approx (20,000 \text{ km})/N$ where the circumference of the Earth $2\pi a_e$ has been approximated by 40,000 km.

Spectral techniques are not only highly accurate, but also nicely sidestep the pole problem faced by grid-point models formulated in terms of latitude and longitude on the sphere. Discretization into spherical harmonics produces no special difficulties at the poles.

A disadvantage of spectral schemes is that fields with strong variation across small-spatial scales, such as most trace gas concentrations (including water vapor),

precipitation, or topography cannot be represented without introducing significant spurious nonlocal oscillations in these fields. This behavior, known as the Gibbs phenomenon is a simple consequence of attempting to represent highly localized features with global basis functions. The nature of these truncation errors is such that the amplitude of spurious oscillations decreases slowly with resolution. The presence of Gibbs oscillations can lead to serious problems in global simulations, such as the formation of negative trace gas concentrations.

Due to the difficulties in spectrally representing fields with intense spatial variability, grid-point models have not been abandoned. In addition, grid-point models can be made more efficient than spectral models at very high resolution. So, while most operational forecasting centers currently use some form of spectral dynamical core in their NWP models, this may change in the next decade. Current research is focused on developing grid-point or finite element approaches on nonstandard grids such as the icosahedral, that is, “bucky-ball” or “soccer-ball,” grid to bypass the pole problem encountered in latitude-longitude discretizations (see section on “Future Directions”) (Table 2).

Parameterizations

Representation of physical processes such as radiation, turbulence, gravity wave drag, convection, and precipitation also became more sophisticated in NWP models as they evolved. The earliest successes in NWP using a filtered barotropic model did not even formally incorporate temperature as a prognostic variable. Today, NWP models may track several condensed water species, as well radiatively active trace gases such as ozone. It was the incorporation of radiative transfer schemes and simple moist convective adjustment schemes into NWP models in the 1960s [35] that led to the first climate models capable of self-consistently representing the basic feature of Earth’s atmospheric general circulation.

Parameterization development for NWP models has followed that for climate models. The suite of parameterizations currently used in global NWP models is identical to that used for climate simulations. It is arguable whether most physics parameterizations

Weather Prediction Models. Table 2 Parameterizations in Weather Forecasting models. The first column gives the usual designation used in the meteorological community. The second column summarizes the effects of the parameterization. The third column indicates the primitive equation forcing term in which tendencies from the parameterization appear

Parameterization	Effects	Included in
Deep convection	Transports heat, moisture and momentum vertically. Damps convergence	H_{phys} , F_{phys} , $C_{i,phys}$
Orographic gravity wave drag	Decelerates flow over mountains	F_{phys}
Planetary boundary layer (PBL) turbulence	Transports heat, moisture and momentum vertically	H_{phys} , F_{phys} , $C_{i,phys}$
Radiation (Solar and IR)	Calculates heating due radiative flux convergence	H_{phys}
Diagnostic cloud	Estimates cloud cover and thickness	
Prognostic cloud	Calculates cloud condensate concentrations and estimates cloud optical properties	H_{phys} , $C_{i,phys}$

exert an appreciable effect on short-term (1–3 day) forecasts. Deep convection and orographic wave drag parameterizations have been shown to exert a significant short-term effect [47–49]. At medium range (3 days to 2 weeks), physical parameterizations are thought to have an important effect on forecast skill (Bengtsson 2000). In addition climate biases resulting from deficient parameterizations can have a negative impact on data assimilation schemes (see section on “Data Assimilation Systems”). This can indirectly affect short-term forecasts by introducing errors in the initial conditions.

How Are NWP Models (Versus Climate Models) Evaluated?

Perhaps the most significant differences between global climate models and NWP models arise from the

different jobs they are expected to perform. Ideally, a solver for the primitive equations, coupled to set of physically realistic parameterizations of processes like convection, should perform both long-term and short-term simulations of the atmospheric flow with equal accuracy. However, this is not the case. Even when run at comparable resolutions, models developed for NWP and climate do not perform each other’s tasks with comparable skill. The likely cause of this discordance is the process of model tuning. The need for “tuning” is widely recognized in both the NWP community and the climate modeling community as an inevitable consequence of using imperfect models [41]. The bulk of model tuning occurs in choosing empirical factors that regulate physical parameterizations. Some of this can be explained (or excused) as an attempt to represent unknown sub-grid distributions of quantities such as water vapor, temperature, or topographic roughness. Unfortunately, the process of selecting these empirical factors can be ad hoc, and parochial in character. Even if optimal tunings exist that combine high accuracy in short-term forecasting as well as unbiased climate simulations, they are unlikely to be found. Groups involved in developing climate models rarely have the computational resources to perform extensive testing of their models at high resolution in forecast mode, while NWP groups are typically under intense operational pressures, and have little time or incentive to examine their models in free-running climate simulations.

There is as yet no set of universally accepted metrics for climate models, although developments in this direction are taking place [50]. Metrics typically targeted by climate modelers include seasonal mean distributions of precipitation and seasonal mean planetary wave patterns. Other, functional, constraints exist for atmospheric models used in climate research. For example, when used in coupled climate simulations, that is, connected through boundary fluxes to ocean and land surface models, obtaining correct global budgets of energy and momentum in an atmospheric model is critical. Thus, intensive tuning of cloud parameterizations is usually conducted to ensure that seasonal and annual-mean radiation budgets at the top of the atmosphere are realistic, to minimize spurious long-term drift in extended simulations. Generally speaking, exact conservation of energy and momentum is a key concern in the design of atmospheric models

for climate, while being of secondary importance in the design of NWP models.

On the other hand, operational NWP models are regularly subjected to a number of rigorous, and more-or-less universally accepted tests at each step in their development. Not surprisingly, these tests emphasize short-term simulation accuracy rather seasonal or annual-mean performance.

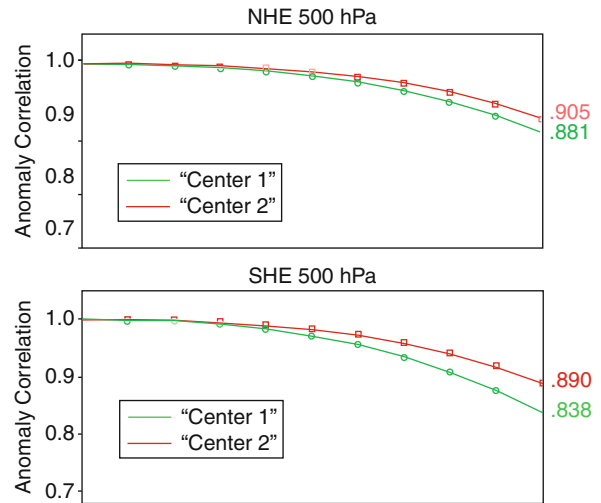
500 hPa Height Anomaly Correlation

At many centers including the European Center for Medium Range Weather Forecasting (ECMWF), the most important measure of global forecast model performance or skill is the 500 hPa height anomaly correlation. This measure is essentially the pattern correlation of two maps of geopotential height anomalies ϕ' interpolated to a pressure level of 500 hPa (corresponding to altitude close to 5 km). A height anomaly is defined the deviation in height from its average value along a latitude circle. One map is an analysis of height anomaly at 500 hPa $\phi'_{ana}(x, y, p_{500}, t_a)$ (see section on “[Initialization and Data Assimilation](#)”) and the second is a map of forecast height anomaly valid at the same time $\phi'_f(x, y, p_{500}, t_a)$ where $t_a = t_i + \Delta t_f$. Here Δt_f is the forecast lead time and t_i is the initiation time. These two height fields are then used to form a correlation.

$$r_{500}(\Delta t_f) = \frac{\langle \phi'_f(t_i + \Delta t_f), \phi'_{ana} \rangle}{\sqrt{\langle \phi'_f, \phi'_f \rangle \langle \phi'_{ana}, \phi'_{ana} \rangle}}$$

where $\langle \rangle$ represents the spatial covariance over some region, typically the southern hemisphere or the northern hemisphere.

Figure 2 shows the average evolution of r_{500} in two major global forecast models for the period January 1 through March 31, 2009. The plots illustrate the state-of-the-art in NWP as of this writing. As expected the pattern correlations decrease with time, but remain quite high, above 0.8, out to a forecast lead of 5 days. Northern hemisphere correlations are higher in both systems, probably reflecting the higher density of in situ measurements available there. Many operational forecasting centers do not allow changes to their systems that degrade this measure of performance.



Weather Prediction Models. Figure 2

Mean evolution of 500 hPa anomaly height correlation in two major global NWP models for the January through March

Although using r_{500} as the single measure of forecast accuracy may seem somewhat restrictive, it should be noted that the geopotential height ϕ at 500 hPa is an integrated measure of the temperature in a deep layer, from the surface to around 5,000 m. So, r_{500} is a concise summary of model performance in a horizontally extensive and deep atmospheric slab.

Skill Scores

The S1 skill score [51, 52] has been used since the 1950s by forecasters at the National Meteorological Center (NMC), and after 1995 the National Center for Environmental Prediction (NCEP), to evaluate forecast performance. An S1 score can be defined for any quantity. It is calculated:

$$S1(\chi) = 100 \times \frac{\int_A |\nabla(\chi_f - \chi_o)| dA}{\int_A \max(|\nabla\chi_f|, |\nabla\chi_o|) dA}$$

where χ_f is forecast χ and χ_o is the observed value for verification. S1 then is the ratio of the integrated absolute gradient in forecast error, normalized by the integrated absolute gradient of the quantity itself, where at each location the larger of the forecast or observed values is used. A value of $S1 = 0$ represents a perfect forecast. The quantity χ used to calculate S1 is typically sea-level pressure or geopotential height.

The S1 skill score was selected by NMC from many measures of forecast quality with guidance from practicing forecasters. Forecasters in 1950s noted that values of S1 around 20 corresponded to very good forecasts, while values of 70 or more represented nearly worthless forecasts. As a result, it became common practice to express “skill” as $2(70 - S1)$, so that now a very good forecast $S1 = 20$ has a skill score of 100, while useless forecasts have a skill score of 0 [52].

Equitable Threat Scores

Evaluation of precipitation forecasts is difficult for a number of reasons. Precipitation is a field with high variance and sharp boundaries. In many circumstances, the important forecast parameter is whether rain or precipitation (above a certain threshold) has occurred. Such categorical forecasts are evaluated using various methods based on matrices of possible outcomes, for example, **YY** – rain is forecast and occurs, **YN** – rain is forecast but does not occur, **NY** – rain is not forecast but occurs, and **NN** – rain is not forecast and does not occur. The most commonly used method is that of Equitable Threat Scores (ETS) [53] which attempts to account for the long-term statistical probability of each category.

Initialization and Data Assimilation

A moment’s reflection shows that establishing initial conditions for a global model of the atmosphere is a nontrivial task. First of all, there may be instrumental errors in the measurements of wind, temperature, humidity, or other quantities needed to specify the initial state of the atmosphere. While perhaps the most obvious problem, instrumental error may also be the least important problem faced in initialization. A more serious problem is hinted at in Fig. 3 which shows the current global distribution of radiosonde balloon launch sights. Radiosondes provide very accurate and reliable measurements of winds, temperatures, and humidities from the surface to around 10 km altitude. Launches are made by international agreement at either 0Z or 12Z (“Z” refers to Greenwich mean time) or at both times, depending on the station. However, as the figure shows, radiosonde launch sites are distributed in a completely unstructured way across the globe.

High concentrations occur in the developed world, with sparse or no coverage over oceans and over less-developed land areas. Even in the developed world the location of sites is determined by human factors – and is more-or-less spatially random. Numerical models of the atmosphere, both spectral and grid point, require a spatially structured set of numbers to begin their integrations. Satellite data presents similar challenges. While it is structured – along orbital tracks – the structure does not conform to the needs of NWP models. In addition, satellite data is asynchronous, that is, sampling occurs continuously as the satellite travels, not at a specified time as with radio-sonde data.

Thus a major challenge faced in NWP is to derive a complete model state on a structured set of points at a single instant in time, from data that may be neither spatially structured nor representative of a single point in time. Naive interpolation in space and/or time is inadequate. The model state derived from the data must not only cover the globe, it must do so while also satisfying a number of other dynamical and physical constraints.

Fluid flows, and vector fields in general, can be decomposed into a sum of divergent ($\nabla \cdot \mathbf{u}_1 \neq 0$) and nondivergent ($\nabla \cdot \mathbf{u}_2 = 0$) components. Generally speaking, the atmospheric motions of most significance in 1–5 day forecasts are characterized by “small” horizontal convergence and divergence in a relative sense. In these flows, two or three of the terms in the momentum equation form a dominant steady-state balance that describes the flow to first-order. These balanced flows are *almost* nondivergent. However, their time evolution can be profoundly affected by the small divergent component. The earliest and most basic balance identified by meteorologists is the so-called geostrophic balance, described below. The subtleties of the divergent wind field in geostrophically balanced flow are what doomed L. F. Richardson’s pioneering NWP experiment.

Geostrophically Balanced Flow

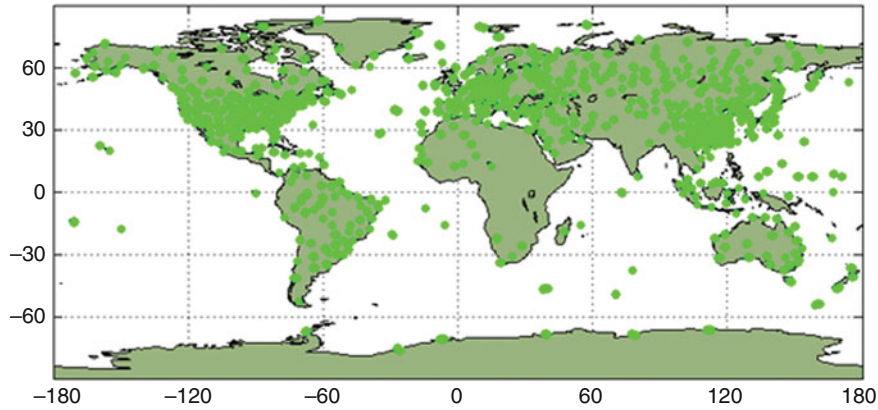
The origin of geostrophic balance is most easily seen by performing a scale analysis of the momentum equation in the primitive equation system. Scale analysis is a common procedure in fluid mechanics to

03May2010,12ZRadiosonde temperatures: 27614 observations

all lat; all lon; all lev; kt = 44; kx = 120; all qcx; all qch

d520_fp.ana.obs.20100503_12z.ods

Observation locations

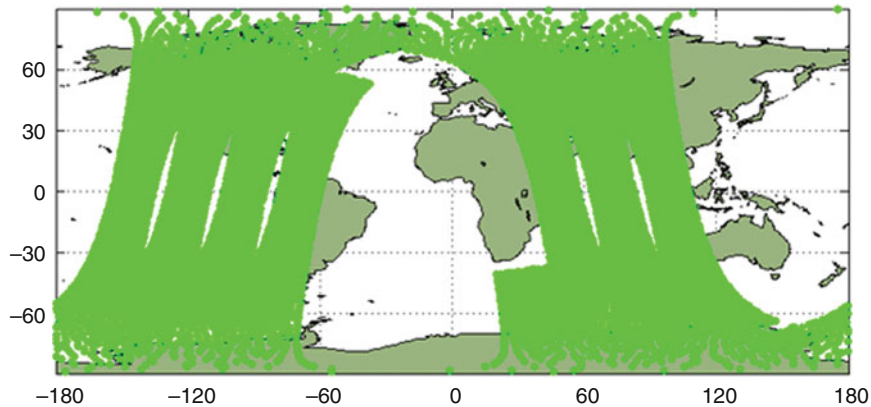


03May2010, 12Z NOAA-15 AMSUA brightness temperature: 178110 observations

all lat: all lon; all lev; kt = 40; kx = 315; all qcx; all qch

d520_fp.ana.obs.20100503_12z.ods

Observation locations



Weather Prediction Models. Figure 3

Locations of radiosonde observations (*top*) and satellite temperature observations (*bottom*) for May 12, 2008 from NASA's GEOS-5 DAS

systematically identify the most important terms in the complicated equations describing fluid flow [4]. It begins by identifying scales of motion for the phenomenon of interest. For midlatitude weather systems, especially after considering the spacing and resolution of radiosonde and satellite data, a reasonable choice of spatial scale L is around 1,000 km. Other reasonable choices of scales are for horizontal wind $U \sim 10 \text{ ms}^{-1}$, and for pressure disturbance $P \sim 10 \text{ hPa}$ or 1,000 Pa. These scales along with local apparent rotation rate f with values $\sim 10^{-4} \text{ s}^{-1}$ in midlatitudes are used to

estimate the sizes of the terms in the equation. For example, $\frac{d}{dt} V$ will be $\sim U^2/L$ where the horizontal advective time scale has been used as the relevant time scale. It is easy to see that the ratio of this term to the Coriolis term $f\mathbf{k} \times V$ will scale as U/fL , which is a key nondimensional parameter in dynamic meteorology known as the Rossby number or Ro . For the scales of motion typical for midlatitude systems, Ro is close to 0.1. So the advective term is likely to be small compared to the Coriolis term and also, it turns out, compared to the pressure gradient.

Thus, the “leading order” balance in midlatitude systems is left as:

$$f\mathbf{k} \times \mathbf{V} = -\alpha \nabla_z p$$

where we have used height as the vertical coordinate. Equivalent expressions exist for other vertical coordinates. In component form this balance is written as:

$$-fv = -\alpha \partial_x p; \quad fu = -\alpha \partial_y p$$

This balance is the dominant feature of midlatitude flow. It is also responsible for one of the most counterintuitive aspects of weather maps in midlatitudes – that the wind blows along pressure contours rather than from high pressure to low pressure. It is also the reason air flows in a counterclockwise sense around low-pressure centers in the northern hemisphere. A perplexing aspect of this balance is that it is a steady-state relation. In other words, the largest forces in the system give no information about its time evolution.

It is easy to see that this leading-order geostrophic flow is horizontally nondivergent. However, spurious divergent flow features can easily appear when constructing initial conditions from observations of horizontal. An idea of the difficulty of this challenge can be obtained by considering the following argument. The relative vorticity of the horizontal wind is determined from:

$$\zeta = \partial_x v - \partial_y u$$

and, further scale analysis of the equations of motion in midlatitudes (see [45]) shows that the ratio $\nabla \cdot \mathbf{V}$ to ζ will typically be close to Ro or ~ 0.1 . The individual horizontal derivative terms in the expressions for vorticity and divergence are of the same order. The small relative magnitude of $\nabla \cdot \mathbf{V}$ is only possible through near cancelation of its much larger component terms. It is not trivial to maintain this cancelation during the data assimilation process. Errors in the divergence can have large effects on surface pressure tendencies since

$$\partial_t p_s \sim p_s \nabla \cdot \mathbf{V}.$$

More complete and correct balance relationships than geostrophic balance can be derived [54]. Balanced initial conditions have been sought in a number of ways during the history of NWP. The most successful of these was perhaps nonlinear normal mode initialization introduced in the 1970s [55–57]. In this technique,

fast and slow normal modes of the nonlinear equations of motion are found through an iterative procedure.

Data Assimilation Systems

Modern operational centers handle the problem of initializing their forecasts by using a data assimilation system (DAS). A central feature of a DAS is the forecast model itself (or a linearized version thereof). Analyses at operational forecasting centers are typically performed four times daily at 00Z, 06Z, 12Z, and 18Z. The analysis procedure combines measurements with a short model forecast, typically 6 h initialized with the previous analysis. This forecast is often referred to as the “first guess” or “background.” The job of the analysis algorithm, denoted by the yellow diamond in Fig. 4, is to blend the myriad sources of data, which includes radiosonde observations satellite measurements, pilot reports, surface station reports, ship buoy measurements, and more, with the forecast to produce an optimal estimate of the state of the atmosphere on the model grid. The process illustrated in Fig. 4 depicts a “3D-Var” system. In 3D-Var, data gathered within an analysis time window, typically 3 h before and after the standard analysis times, is assumed to be synchronous. The analysis then consists of an optimal blending of measurements and forecast background in space.

This optimal blending can be expressed as a “cost function minimization.” The cost function is written in matrix form [2, 54, 58]:

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + [\mathbf{y} - H(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - H(\mathbf{x})] + J_b(\mathbf{x})$$

where $\mathbf{x}$ is a model “state vector,” that is V , T , q and possibly other analyzed species on the model’s numerical grid (or spectral decomposition). The quantity $\mathbf{x}_b$ is the model state from a forecast, that is, the first guess or background, and the matrix $\mathbf{B}$ is the background, or forecast, error covariance matrix. This matrix is a key piece of the analysis algorithm, and is estimated by examining differences between forecasts radiosonde observations [59] or more recently by calculating the covariance of different short forecasts, for example, 24 h and 48 h, valid at the same time [60, 61].

The second term in the cost function contains the observation error covariance matrix $\mathbf{R}$, the observation

vector $\mathbf{y}$ and a vector $H(\mathbf{x})$ which is the result of the “observation operator” H acting on the model state $\mathbf{x}$. In the case of an observation taken by a thermometer placed at a model grid-point, the observation operator would simply select the appropriate element of $\mathbf{x}$. However, in the case of remote satellite observations, which directly measure radiances (photons) from the atmosphere, H could represent a complex radiative transfer calculation using for example T and q from the model state to estimate the radiance measured by a particular instrument. The approach of transforming model quantities into a form that is directly comparable with observations, sometimes referred to as “radiance assimilation,” led to dramatic increases in the positive impact of satellite measurements on forecasting [62].

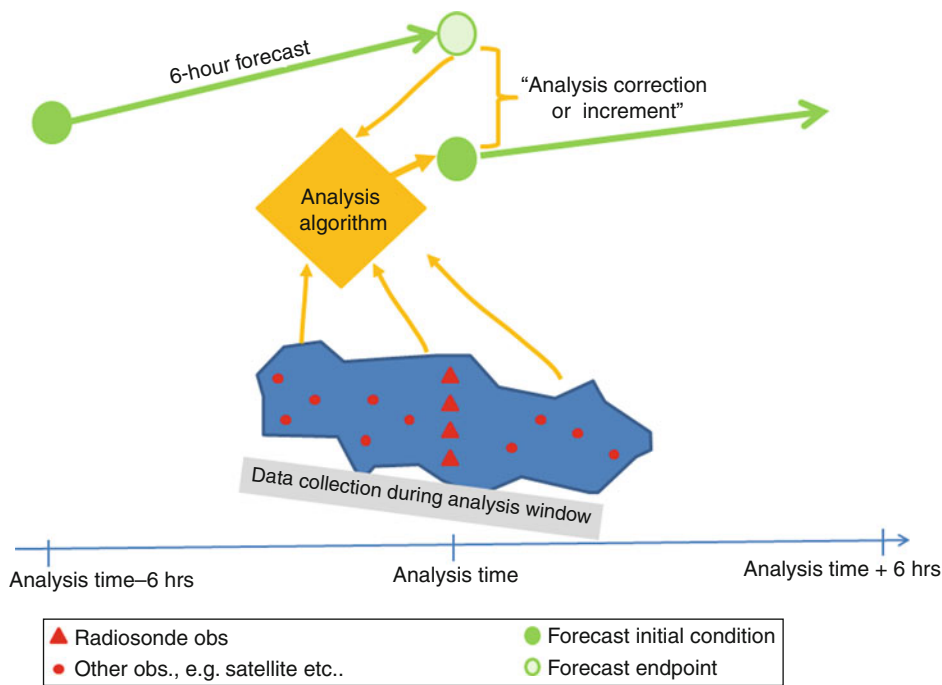
The third term in the cost function represents balance constraints on the flow, such as those discussed in the section on “Geostrophically Balanced Flows”. Inclusion of such a term at NCEP has eliminated the need for a separate initialization procedure for forecasts [60].

The task of the analysis algorithm is to find the model state $\mathbf{x}_a$ that minimizes the cost function $J(\mathbf{x})$.

For more details on how this solution is actually accomplished the reader is referred to discussion in Chap. 5 of Kalnay (2003) [54]. 3D-var as described here is used by NCEP as well as in slightly modified version by NASA’s Global Modeling and Assimilation Office (GMAO) in their GEOS-5 DAS [63]. A somewhat different approach known as 4D-Var is used at ECMWF. This approach takes into account the possibly asynchronous nature of data when formulating the cost function. For more details, see the discussion in Kalnay (2003).

Ensemble Forecasting

The early thinking of researchers in NWP was that the central problem of forecast initialization was to correctly filter out rapid divergent motions, and, that once this was accomplished no fundamental limits on atmospheric predictability existed. A series of seminal papers by Edward N. Lorenz proved this to be incorrect [64–66]. Lorenz showed that simple analogs to atmospheric equations of motion possess a sensitive dependence to



Weather Prediction Models. Figure 4
Schematic diagram of 3D-var analysis

initial conditions. In other words, small differences in even well-balanced initial conditions will cause forecasts to become uncorrelated after a finite time.

Modern operational forecasting centers typically perform ensembles of many runs with slightly different initial conditions, as well as a single higher-resolution “deterministic” run to produce forecasts for a given time [54, 67]. The generation of ensemble members is a nontrivial task. Ideally, the members of the ensemble should vary in special directions in phase space that are related to the most rapidly growing instabilities in the flow [54, 68].

Future Directions

Weather forecasts have improved demonstrably during the 50 year history of NWP. This is illustrated in Fig. 5, which shows the evolution of r_{500} at 3, 5, and 7 days over the last 30 years in the ECMWF system [69]. Part of the improvement is traceable to the explosion in the amount of satellite data over the last 30 years. However, a large part of the improvement in skill is due to improvements in the forecast and analysis “system,” such as increased forecast model resolution, improved analysis algorithms etc. This is nicely demonstrated in the bottom panel of Fig. 5, which shows the evolution of skill in retrospective forecasts, using the *current* ECMWF system on the historical data base. The skill of retrospective forecasts is significantly higher, indicating that the improved forecast and analysis system makes a significant, perhaps the dominant, contribution to the overall increase in skill seen in the last 30 years. One aspect of improvement that is clearly due to improved data sources (satellites) is the convergence of skill in southern and northern hemispheres.

Scalable Dynamical Cores

As computing power increases NWP model resolution also continues to increase. In the last several years, the increase in computer power has appeared primarily in the form of massively parallel machines with larger and larger numbers of processors, rather than in the form of faster individual processors. This means that “time-to-solution” has not decreased dramatically in recent years, but the size of feasible calculations has increased dramatically. This has stimulated the development of scalable models. Scalability means that

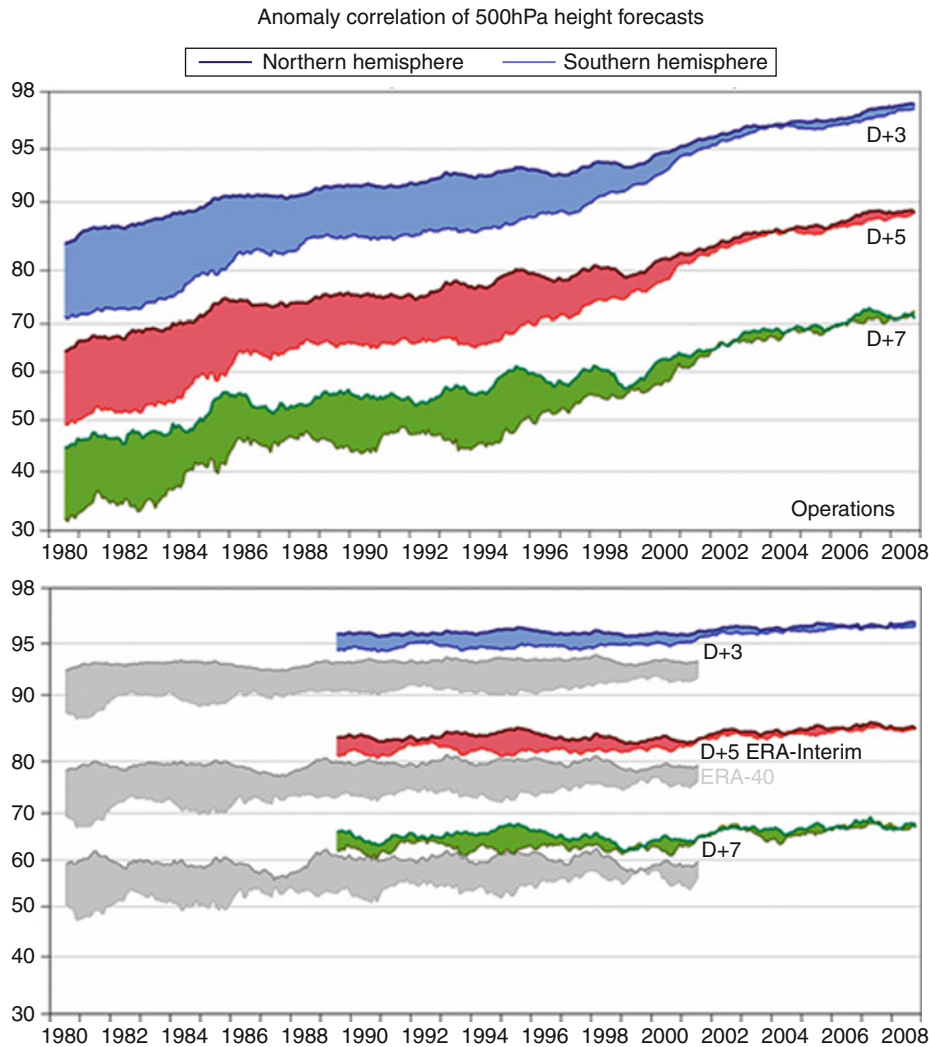
model speed increases more-or-less linearly with the number of processors used. A trivial example of perfectly scalable problem is the addition of 1,000 pairs of numbers $a + b = c$. If one processor is available then the calculation will take 1,000 CPU time units. If 1,000 processors are available the entire calculation can take place in 1 CPU time unit. However, communication between processors also costs time. In any real numerical model of the atmosphere processors eventually need information residing in other processors. This prevents numerical models of the atmosphere from scaling perfectly.

The amount of cross-processor communication required can vary widely depending on model design. Ideally decomposition will maximize the ratio of area to perimeter in computational subdomains to minimize the need for cross-processor communication. The scalability of grid-point models on latitude-longitude grids is severely hampered by the need to apply polar filters to overcome numerical instabilities that arise due to the convergence of longitude lines at both poles. These filters typically require knowledge of atmospheric fields all the way around latitude circles. Thus, modelers must either decompose the globe into thin computational domains circling the globe, which will lead to large communication requirements in the north–south direction, or they must pay the cost of frequent “gathers” to obtain the necessary inputs for polar filters. Transforms in spectral models likewise require knowledge of fields around latitude circles.

Recent efforts in numerical techniques for global atmospheric models have focused on the development of grid-point or finite element models on nonstandard grids [70–73]. Several examples of such grids are shown in Fig. 6. These grids have fairly uniform grid cell sizes over the entire globe. Polar filters are therefore not required.

Nonhydrostatic Dynamics

Once horizontal resolution becomes much finer than 10 km, nonhydrostatic effects must be taken into account. This will require models based on a different set of equations. One option is to simply use the full Euler or Navier Stokes equations [39, 74] and pay the costs associated with the short time-steps required by the presence of acoustic waves. Another approach is to



Weather Prediction Models. Figure 5

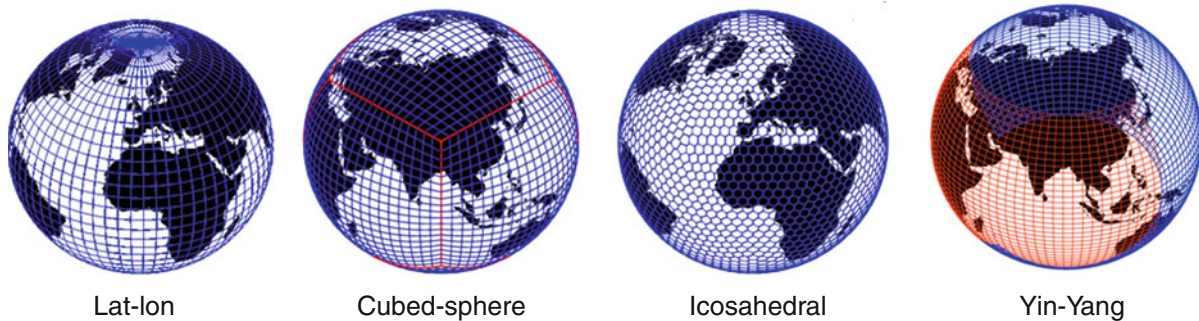
Evolution of forecast skill at ECMWF, adapted and extended from the study of Simmons and Hollingsworth (2002). The *top panel* shows a history of r_{500} at 3, 5, and 7 days from the ECMWF operational forecast beginning in 1980. The *colored and shaded areas* are bounded by southern hemisphere skill below and northern hemisphere skill above. The *lower panel* shows retrospective forecasts produced using two versions of current ECMWF analysis and forecast systems. In these cases, the “system” is fixed in time, while data inputs evolve in actual historical fashion

use an inelastic equations system [75], but this requires solution of an elliptic equation which is an intrinsically nonlocal procedure and again raises cross-processor communication costs.

Seamless Models for Climate and Weather

Modern global NWP and climate models are essentially the same. Both use the same set of dynamical equations

(see section on “[Primitive Equations](#)”). Both also use the same set of physical parameterization schemes (see section on “[Parameterization](#)”). In practice, differences do exist between global models of the atmosphere intended for climate simulation and those intended for forecasting. Operational global NWP models have typically used resolutions that are a factor of 8–16 times higher than those used in long climate simulations. Other differences arise, more-or-less unintentionally,



Weather Prediction Models. Figure 6

The standard latitude-longitude “lat-lon” grid (*leftmost globe*) compared with newer, non-traditional grids for global atmospheric models

in the tuning process as a consequence of the different metrics used in the evaluation of NWP and climate models (see section on [How are NWP models \(versus climate models\) evaluated?](#)).

Climate researchers understand that atmospheric phenomena such as squall lines and tropical cyclones may play a role in establishing climate on both regional and global scales. Such “mesoscale” features are not resolved in climate models with resolutions of 100 km or coarser. However, the continued increase in computer power, and the recent emphasis on massively parallel architecture, will allow decadal or even century-long simulations at resolutions close to 10 km in the near future. At these resolutions, mesoscale circulations should be well represented. These resolutions will also present novel challenges to the sub-grid parameterizations used in climate models, as assumptions about scale-separation and statistical equilibrium become questionable.

Bibliography

1. Tribbia J, Anthes R (1987) Scientific basis of modern weather prediction. *Science* 237:493–499
2. Lynch P (2006) The emergence of numerical weather prediction: Richardson’s dream. Cambridge University Press, Cambridge, MA
3. Richardson L (1922) Weather prediction by numerical process. Cambridge University Press, Cambridge
4. Gill A (1982) Atmosphere–ocean dynamics. Academic, London
5. Simpson J (1994) Sea breeze and local winds. Cambridge University Press, New York
6. Lau K, Li M (1984) The monsoon of East Asia and its global associations—survey. *Bull Amer Meteor Soc* 65:114–125
7. Webster P, Tomas R (1998) Monsoons- processes, predictability, and the prospects for prediction. *J Geophys Res* 103(C7): 14451–14510
8. Douglas M, Maddox R, Howard K, Reyes S (1993) The Mexican monsoon. *J Clim* 6(8):1665–1677
9. Koch S, Handley C (1997) Operational forecasting and detection of mesoscale gravity waves. *Weather Forecast* 12(2)
10. Moncrieff M, Liu C (1999) Convection initiation by density currents: role of convergence, shear, and dynamical organization. *Mon Weather Rev* 127(10):2455–2464
11. Lac C, Lafore J, Redelsperger J (2002) Role of gravity waves in triggering deep convection during TOGA COARE. *J Atmos Sci* 59(8)
12. Houze R Jr (2004) Mesoscale convective systems. *Rev Geophys* 42. doi:10.1029/2004RG000150
13. Grabowski W, Moncrieff M (2006) Large-scale organization of tropical convection in two-dimensional explicit numerical simulations. *Q J Roy Meteor Soc* 127(572):445–468
14. Smith R (1979) The influence of mountains on the atmosphere. *Adv Geophys* 21:87–230
15. Longuet-Higgins M (1964) Planetary waves on a rotating sphere. *Proc R Soc Lond A Math Phys Sci* 279(1379):446–473
16. Andrews D, McIntyre M (1976) Planetary waves in horizontal and vertical shear: the generalized Eliassen-palm relation and the mean zonal acceleration. *J Atmos Sci* 33:2031–2048
17. Hoskins B, Ambrizzi T (1993) Rossby wave propagation on a realistic longitudinally varying flow. *J Atmos Sci* 50: 1661–1671
18. Charney J (1947) The dynamics of long waves in a baroclinic westerly current. *J Meteor* 4(5):135–162
19. Eady E (1949) Long waves and cyclone waves. *Tellus* 1(3):33–52
20. Kiladis G, Wheeler M, Haertel P, Straub K, Roundy P (2009) Convectively coupled equatorial waves. *Rev Geophys* 47. doi:10.1029/2008RG000266
21. Burpee R (1972) The origin and structure of easterly waves in the lower troposphere of North Africa. *J Atmos Sci* 29(1):77–90
22. Landsea C et al (1993) A climatology of intense (or major) Atlantic hurricanes. *Mon Weather Rev* 121(6):1703–1713

23. Pytharoulis I, Thorncroft C (1999) The low-level structure of African easterly waves in 1995. *Mon Weather Rev* 127(10)
24. Thorncroft C, Hodges K (2001) African easterly wave variability and its relationship to Atlantic tropical cyclone activity. *J Clim* 14(6):1166–1179
25. Madden R, Julian P (1994) Observations of the 40–50-day tropical oscillation: a review. *Mon Weather Rev* 122(5):814–837
26. Zhang C (2005) Madden-Julian oscillation. *Rev Geophys* 43(2). doi:10.1029/2004RG000158
27. Liebmann B, Hendon H, Glick J (1994) The relationship between tropical cyclones of the western Pacific and Indian Oceans and the Madden–Julian oscillation. *J Meteor Soc Japan* 72(41):1–412
28. Maloney E, Hartmann D (2000) Modulation of hurricane activity in the Gulf of Mexico by the Madden-Julian oscillation. *Science* 287(5460):2002
29. Maloney E, Hartmann D (2000) Modulation of eastern North Pacific hurricanes by the Madden-Julian oscillation. *J Clim* 13(9)
30. Elsner J, Jagger T, Niu X et al (2000) Changes in the rates of North Atlantic major hurricane activity during the 20th century. *Geophys Res Lett* 27(12):1743–1746
31. Hall J, Matthews A, Karoly D (2001) The modulation of tropical cyclone activity in the Australian region by the Madden–Julian oscillation. *Mon Weather Rev* 129:12
32. Bond N, Vecchi G (2003) The influence of the Madden–Julian oscillation on precipitation in Oregon and Washington. *Weather Forecast* 18(4):600–613
33. Jones C, Waliser D, Lau K, Stern W (2004) The Madden–Julian oscillation and its impact on northern hemisphere weather predictability. *Mon Weather Rev* 132(6)
34. Waliser DE, Lau KM, Stern W, Jones C (2003) Potential predictability of the Madden–Julian oscillation. *Bull Amer Meteor Soc* 84:33–50
35. Manabe S, Smagorinsky J, Strickler R (1965) Simulated climatology of a general circulation model with a hydrologic cycle. *Mon Weather Rev* 93(12):769–798
36. Namias J (1980) The early influence of the Bergen School on synoptic meteorology in the United States. *Pure App Geophys* 119(3):491–500
37. Grenas S (2005) Vilhelm Bjerknes' vision for scientific weather prediction. In: *The Nordic seas: an integrated perspective: oceanography, climatology, biogeochemistry, and modeling*. American Geophysical Union, Washington, DC, p 357
38. Hinkelmann K (1951) Der Mechanismus des meteorologischen Lärmes. *Tellus* 3:285–296
39. Laprise R (1992) The Euler equations of motion with hydrostatic pressure as an independent variable. *Mon Weather Rev* 120(1):197–207
40. Shuman F, Hovermale J (1968) An operational six-layer primitive equation model. *J Appl Meteor* 7(4):525–547
41. Bengtsson L (2001) The development of medium range forecasts. In: *50th anniversary of numerical weather prediction, commemorative symposium*, Potsdam
42. Kasahara A (2000) On the origin of cumulus parameterization for numerical prediction models. *Gen Circ Model Dev* 70:199–224
43. Tiedtke M (1993) Representation of clouds in large-scale models. *Mon Weather Rev* 121:3040–3061
44. Del Genio A, Yao M, Kovari W, Lo K (1996) A prognostic cloud water parameterization for global climate models. *J Clim* 9:270–304
45. Haltiner G, Williams R (1983) *Numerical prediction and dynamic meteorology*, 2nd edn. Wiley, New York
46. Leith C (1965) Numerical simulation of the earth's atmosphere. In: *Methods in computational physics*, vol 4. Academic, New York
47. Webster S, Brown A, Cameron D, Jones C (2006) Improvements to the representation of orography in the Met Office Unified Model. *Q J Roy Meteor Soc* 129(591):1989–2010
48. Clark A, Gallus W Jr, Xue M, Kong F (2009) A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles. *Weather Forecast* 24(4):1121–1140
49. Bhanu Kumar O, Ramalingeswara Rao S, Muni Krishna K (2009) Role of cumulus parameterization schemes in simulating heavy rainfall episodes off the coast of Maharashtra state during 28 June–4 July 2007. *Meteorol Atmos Phys* 105(3):167–179
50. Gleckler P, Taylor K, Doutriaux C (2008) Performance metrics for climate models. *J Geophys Res* 113:D06104
51. Teweles S, Wobus H (1954) Verification of prognostic charts. *Bull Amer Meteor Soc* 35:455–463
52. Shuman F (1989) History of numerical weather prediction at the National Meteorological Center. *Weather Forecast* 4:286–296
53. Gandin L, Murphy A (1992) Equitable skill scores for categorical forecasts. *Mon Weather Rev* 120:361–370
54. Kalnay E (2003) *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press, Cambridge, MA
55. Machenhauer B (1977) On the dynamics of gravity oscillations in a shallow water model, with applications to normal mode initialization. *Beitr Phys Atmos* 50:253–271
56. Baer F, Tribbia J (1977) On complete filtering of gravity modes through nonlinear initialization. *Mon Weather Rev* 105(12):1536–1539
57. Tribbia J (1984) A simple scheme for high-order nonlinear normal mode initialization. *Mon Weather Rev* 112:278–284
58. Daley R (1993) *Atmospheric data analysis*. Cambridge University Press, New York
59. Thiebaut HJ, Pedder MA (1987) Spatial objective analysis with applications in atmospheric science. Academic, London/Orlando, p 308
60. Parrish D, Derber J (1992) The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon Weather Rev* 120(8):1747–1763
61. Buehner M, Gauthier P, Liu Z (2005) Evaluation of new estimates of background-and observation-error covariances for variational assimilation. *Q J Roy Meteor Soc* 131(613):3373–3384
62. Derber J, Wu W (1998) The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. *Mon Weather Rev* 126(8):2287–2299
63. Rienecker M, Suarez M, Todling R, Bacmeister J, Takacs L, Liu H, Gu W, Sienkiewicz M, Koster R, Gelaro R et al (2008) The GEOS-5

data assimilation system-documentation of versions 5.0.1, 5.1.0. NASA Technical Memorandum, NASA/TM-2007, 104606

64. Lorenz E (1965) A study of the predictability of a 28-variable atmospheric model. *Tellus* 17(3):321–333
65. Lorenz E (1969) The predictability of a flow which possesses many scales of motion. *Tellus* 21(3):289–307
66. Lorenz E (1969) Atmospheric predictability as revealed by naturally occurring analogues. *J Atmos Sci* 26(4):636–646
67. Lewis J (2005) Roots of ensemble forecasting. *Mon Weather Rev* 133(7):1865–1885
68. Toth Z, Kalnay E (1993) Ensemble forecasting at NMC: the generation of perturbations. *Bull Amer Meteor Soc* 74:2317–2330
69. Simmons A, Hollingsworth A (2002) Some aspects of the improvement in skill of numerical weather prediction. *Q J Roy Meteor Soc* 128(580):647–678
70. Ringler T, Heikes R, Randall D (2000) Modeling the atmospheric general circulation using a spherical geodesic grid: a new class of dynamical cores. *Mon Weather Rev* 128(7):2471–2490
71. Skamarock W, Klemp J, Ringler T, Thuburn J (2008) A hexagonal C-grid atmospheric core formulation for multiscale simulation on the sphere. AGU fall meeting abstracts, San Francisco, CA, p 01
72. Putman W, Lin S (2009) A finite-volume dynamical core on the cubed-sphere grid. *Astron Soc Pac Conf Ser* 406:268
73. Taylor M, Fournier A (2010) A compatible and conservative spectral element method on unstructured grids. *J Comput Phys* 229:5879–5895
74. Klemp J, Skamarock W, Dudhia J (2007) Conservative split-explicit time integration methods for the compressible nonhydrostatic equations. *Mon Weather Rev* 135(8):2897–2913
75. Grabowski W, Smolarkiewicz P (2002) A multiscale anelastic model for meteorological research. *Mon Weather Rev* 130(4)

Wide Area Monitoring, Protection and Control

JAMES S. THORB, ARUN G. PHADKE

Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

Article Outline

Glossary
 Definition of the Subject
 Power System Monitoring
 Protection
 Power System Control
 Future Directions
 Bibliography

Glossary

Adaptive protection A protection philosophy which permits and seeks to make adjustments automatically in various protection functions in order to make them more attuned to prevailing system conditions.

Fault A short circuit in the power system.

GPS Global Positioning System.

Hidden failure A defect or error in relay that does not manifest itself immediately but which can cause a miss-operation when the system is stressed.

NERC North American Electric Reliability Corporation.

Phasor A complex number describing a sinusoidal signal.

PMU Phasor Measurement Unit.

RTU Remote Terminal Unit.

SCADA Supervisory Control and Data Acquisition.

Synchrophasor A phasor measurement with a precise time tag.

WAMS Wide Area Measurement System.

Definition of the Subject

The Electric Power Infrastructure

A power system consists of generators which convert energy from some primary form – such as coal, wind, gas, sunlight, or water flow – into electric power. Most generators in modern power systems produce alternating currents and voltages (AC) which are then transmitted over appropriate distances to supply loads. It is economical to transmit power at higher voltages in order to limit losses in the transmission lines. Voltages of desired values are obtained with the help of transformers which let through most of the power (small amount of losses are incurred in the transformers), only changing voltages and currents from one side of the transformer to the other. The “power” handled by the transformers is the product of currents and voltages, so that higher voltages have accompanying smaller currents. In general, the longer the distance for transmission, the greater is the economical voltage at which to transmit the power. Most power transmission is through overground transmission lines, while in special circumstances (such as crossing a large body of water or in congested urban areas) underground cables

are used. There is a range of voltages in use for transmission: 800 (kilovolts – thousands of volts), 500, 400 kV, etc. The household voltages in most countries are 120 V or 240 V which are obtained through stages of transformations as power is passed through transmission, sub-transmission, and distribution systems. For technical reasons, sometimes it is advantageous to convert AC power to DC (direct current) power by high-power electronic converters, transmit it as DC power, and convert it back to AC at the receiving end. The power-handling capacity of total system generation is expected to be some 10–20% greater than the expected load in order to provide a margin of security. Similarly, the transmission system is planned with sufficient number of alternative paths connecting loads to generators so that should some lines get disconnected due to faults, the loads continue to be served. In North America and many countries, the frequency of the AC system is nominally set at 60 Hz, while in some other countries, it is 50 Hz. The actual frequency of the system stays close to the nominal values, and in well-regulated power systems, the variations from nominal frequency are less than 0.1 Hz. The North American power system has approximately 900 gigawatts (900,000,000,000 W) of power generation capacity – consisting of about 70% thermal power generators, 14% hydroelectric generators, 13% nuclear, and small amounts of wind, solar, and other renewable power sources. The continent is crisscrossed by about 200,000 circuit miles of high-voltage transmission lines, further supplemented by lower-voltage distribution circuits serving customers at lower voltages at the load centers. The generation capacity in North America is sufficient to meet load demands with adequate margins overall, although in spot locations, the available margins or transmission capacity may not be sufficient.

Operating the Power Grid Reliably

In order to provide electric power reliably to consumers, it becomes necessary to manage the generation, transmission, and consumption of electric power in an efficient manner. Since the load demand is controlled by the consumers and may change randomly depending upon consumers' wishes, and since the generation must match this demand (and losses) instant by instant, an elaborate system of controls is put in place at

generators and at Energy Management Centers on the power grid. Such controls must take into account the possibility of faults in the system due to natural disasters (floods, hurricanes, and fires), equipment failure, or man-made disasters. These events are contingency events, and the power grid must continue to meet the demands of consumers with minimum interruptions when faced with such contingencies. In addition, the power system must maintain its operating frequency within a narrow band around the nominal value (60 Hz in North America), and the operating voltage at the consumers' locations within a narrow band around its nominal value. To be prepared to deal with a reasonable set of contingencies adequately, it is necessary that the system operators are made aware of the condition of the power infrastructure accurately at all times. This is achieved through a system of measurements made on the power grid at frequent intervals, from which the network flows and voltages are estimated. The technical term for this process is "state estimation" of the power system. The ability of the power system to meet its loads in the presence of contingencies is much enhanced if the state of the system is known precisely and with sufficient frequency so that as the transmission system, loads, and generation change during the day, the changes are tracked accurately by the measurement process. A measurement technique made possible by the advent of digital computer-based measurements and the Global Positioning System (GPS) is the highly precise and effective "synchrophasors" technology.

Sine Waves, Phasors, and Synchrophasors

Alternating currents and voltages vary as sinusoids with a period determined by the frequency of the system. Thus, in North America where the frequency is 60 Hz, a nominal voltage or current will vary with time as shown in Fig. 1. Note that the sinusoid is shown to be a pure sine wave, whereas in practice, the waveform could have some noise in it. A *phasor* is a mathematical way of representing this sinusoid. It is a complex number with its magnitude equal to the RMS (root mean square) value of the sinusoid, and its phase angle is the angle measured with respect to the start of the time axis and the peak of the sinusoid. One period is equivalent to 2π radians or 360° . Angles measured to the left of $t = 0$ are positive angles, and angles measured to the

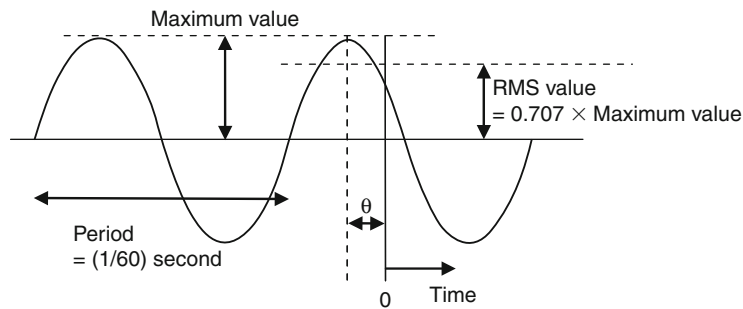
right are negative angles. Thus, the phasor representing the sinusoid in Fig. 1 is given by:

$$\text{Phasor} = Xe^{j\theta} = X \cos \theta + j X \sin \theta$$

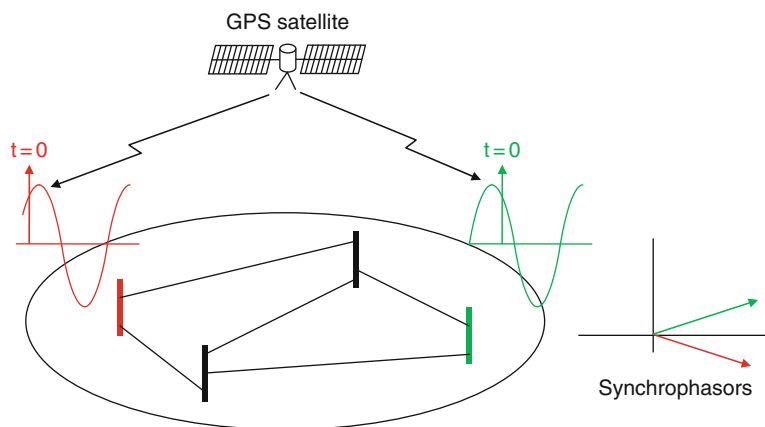
where X is the RMS value of the sinusoid. Phasors are generally calculated from samples taken from a sinusoid over a time span known as data window, which is usually of the order of one period of the sinusoid.

The GPS satellite system transmits (among other data) a 1-pulse-per-second (1 pps) which can be received by GPS receivers. A characteristic of the GPS system is that a 1 pps received by any receiver placed anywhere on earth is within $1 \mu\text{s}$ of the pulse received by another GPS receiver. Consider a power system shown in Fig. 2 in which different power facilities

(substations) are separated by large distances. The sinusoids of voltages and currents in the two substations would in general be different from each other in magnitudes and phase angles of the sinusoids. A sinusoid at the substation shown in red in Fig. 2 is measured from the instant shown as $t = 0$, which is identified with the 1 pps received from the satellite, while the measurement at the substation shown in green has the $t = 0$ time marker at the same instant, and measures the phasor at that location. The two synchrophasors can now be placed on the same phasor diagram and show correctly the relative magnitudes and phase angles of the two signals as they exist at that instant. Synchrophasors can be measured at all network buses and assembled at the system control center to form a coherent picture of the



Wide Area Monitoring, Protection and Control. Figure 1
A sinusoidal voltage or current and its phasor representation



Wide Area Monitoring, Protection and Control. Figure 2
Voltages and currents in substation at different locations

state of the power system. Such a picture can be refreshed as frequently as once per cycle of the power system frequency and is thus able to track the behavior of the power system as it goes through normal or abnormal changes brought about by daily load variations or major disturbance. This ability to monitor the power system continuously with precise simultaneous measurements is made possible by the technology of the synchrophasors.

Importance of Positive Sequence Measurements

The entire AC power system is made up of three-phase sources and loads. (Three-phase system makes the most economical use of the capability of large generators and motors.) The three phases are considered to be balanced in most normal operating conditions. When faults or unusual disturbances occur, the system may become severely unbalanced, but such unbalances last only for a short time, and the system soon returns to a balanced state. It is well to provide a more precise definition of balanced operation.

When the voltages and currents in a three-phase AC system are of equal magnitude and differ in phase from each other by equal amounts ($360^\circ/3 = 120^\circ$), the system is said to be balanced. This is illustrated in Fig. 3, where the three signals (voltages or currents) of the three phases are illustrated by red, green, and blue colors. The three phases are usually labeled.

“a,” “b,” and “c” in most power systems, although in some countries, they are referred to by three different colors (red, yellow, and blue). If the power system is balanced, it turns out that it is not necessary to consider it to be a three-phase system in most computational tasks; all analyses can be performed in a single-phase mode, and from the single-phase results, the conditions

in the three phases can be inferred. This leads to much simpler computational procedures for all power system problems. Technically, this single-phase representation is known as the “positive sequence” representation. In general, a three-phase parameter (voltage, current, or impedance) can be represented by three “sequence” components: positive, negative, and zero. Under balanced conditions, only the positive sequence component exists, and the negative and zero sequences are zero. When severe unbalances occur (e.g., when there is a line to ground fault), negative and zero sequence components become significant. Even in normal operation, there may be a small amount of unbalance present in the power system; thus, small amounts of negative and zero sequence components may exist on the power system during normal operating conditions. However, for most analyses, it is assumed that the unbalance can be neglected, and the power system may be assumed to be balanced; hence, only a single-phase representation using positive sequence voltages, currents, and impedances is appropriate.

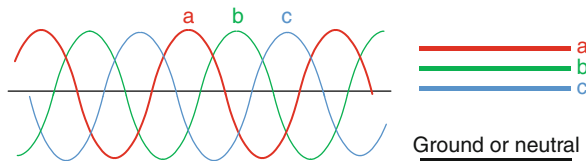
Since most power system problems are handled with positive sequence quantities, it is most advantageous to measure synchrophasors of positive sequence voltages and currents. This is accomplished by measuring synchrophasors of individual phase voltages and currents (X_a , X_b , and X_c) and calculating positive sequence synchrophasors (X_1) from them:

$$X_1 = 1/3 \left[X_a + \left(-1/2 + j \sqrt{3}/2 \right) X_b + \left(-1/2 - j \sqrt{3}/2 \right) X_c \right]$$

Phasor Measurement Unit (PMU)

Synchrophasors are measured by Phasor Measurement Units (PMU) installed in power system substations. A simplified schematic diagram of the PMU is shown in Fig. 4.

Currents and voltages from the power system are transformed to measurement levels by instrument transformers, which are then supplied to the filter stage of the PMU which removes destructive induced voltages and currents as well as higher frequencies in order to avoid aliasing errors in the sampled data. The GPS receiver provides the 1 pps from the GPS satellites and is divided into desired sampling rates by a phase-locked clock. Typical sampling rates may vary between

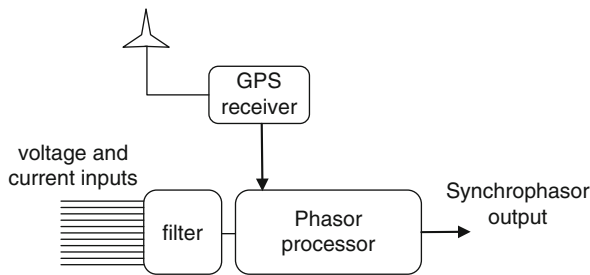


Wide Area Monitoring, Protection and Control. Figure 3 Balanced voltages and currents in a three-phase power system

12 samples per cycle of the power frequency to higher sampling rates measured in kilohertz. The phasor processor converts samples of each phase into a synchrophasor and finally into a positive sequence synchrophasor measurement which is assigned a time stamp before the measurements are transmitted over a suitable communication network to provide a stream of synchrophasor measurements. Typical reporting rates for the synchrophasors are of the order of once every cycle to once every few cycles depending upon the user's requirements.

WAMS Architecture Overview

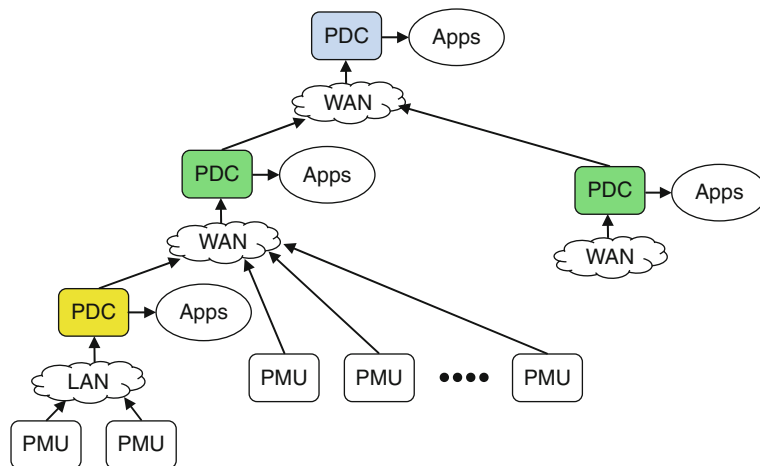
WAMS (Wide Area Measurement System) consists of Phasor Measurement Units (PMU) installed in power system substations and connected to one or more layers



Wide Area Monitoring, Protection and Control. Figure 4
A simplified block diagram of a generic PMU

of Phasor Data Concentrators (PDC) so that system-wide phasor data can be made available at various regional or central locations for engineering applications (see Fig. 5). The PMUs are placed in substations, where they collect voltage and current data. If multiple PMUs exist within one substation, their outputs are streamed to a PDC within the substation – shown in yellow color – over a local area network (LAN). Where there is one PMU per substation, the data from those PMUs and the substation PDC is connected to a regional PDC – shown in green color – over a wide area network (WAN). Regional PDCs in turn can be connected to a central PDC – shown in blue color – over another WAN. Each PDC basically collates data from the lower-level devices and by matching their time stamps creates a data stream as though it was a PMU with many inputs. The PDC may also remove faulty data, deselect some data, or alter the phasor reporting rate as required by specific applications.

The LAN and WAN are likely to be private wired or fiber-optic communication networks on which high communication rates can be achieved. It is not expected that public WAN would be used for this purpose, as the questions of security and communication delays may rule this option out. In many cases where fiber-optic options are not available, digital microwave, leased phone lines, or any other available communication channels may be utilized by the WAMS system. Of course, the throughput rate as well



Wide Area Monitoring, Protection and Control. Figure 5
Wide area measurement system architecture

as data latency of communication media dictates which of the WAMS applications can be successfully implemented in such systems.

Suitable application tasks may be implemented at each PDC. Clearly, the data available at each PDC covers differing regions of the power system. Applications of WAMS for power system monitoring, protection, and control will be described in the following sections, where the amount of data needed by a specific application and the region of the power system which must be visible to the application will be discussed. As a general rule, the applications at lower levels of PDC require lesser amounts of data and respond more quickly than those at higher-level PDCs. In general, it has been found that the fast-responding applications can tolerate measurement latencies in the order of 50–100 ms.

Power System Monitoring

Reconstruction of Events

The job of reconstructing the events involved in a major disturbance is a daunting one. Months and immense effort is involved. It is typically difficult even to determine the sequence of events. Phasor measurements with precise time tags were not numerous in 2003, but the few that did exist stood out among the unsynchronized data. From the simple function of fault recorders to complicated reconstructions involving thousands of buses, time-synchronized measurements are vital.

State Estimation

One of the conclusions of the report on the 1965 Northeast blackout was that, at that point, power system operators had insufficient information about the system they were controlling. State estimation and the modern energy management system were created to deal with these problems. Early state estimation “polled” SCADA systems connected to Remote Terminal Units in substation by moving systematically from substation to substation covering the entire system is the total scan. The measurements were power flows on lines, power inputs (injections) into buses, and occasional voltage magnitudes. It was necessary to assume that the system was *static* during the scan, i.e., the system did not change during the scan. Another view

of the resulting estimate of the state is that it is the state of a hypothetical system which could support the complete set of measurements. Depending on how long the scan takes and the changes in the system that took place during the scan, the hypothetical system may not exist or may be quite different from the real system. The state of the power system was taken as the positive sequence voltages at the buses which could not be measured directly. The measurements and the estimated state were all complex numbers, magnitudes and angles, or real and imaginary parts. The calculation had to solve large numbers of nonlinear equations and was iterative. That combined with the data scan made the process slow with new estimates available every few minutes [1]. The result was a picture of the state of the system presented to operators and to the software that did the energy management (contingency analysis, for example). Faster computers and improved communication have reduced minutes to seconds over the decades.

The complex voltage at bus k with angle θ_k and magnitude V_k is given by

$$E_k = V_k e^{j\theta_k}$$

The measurements are complex flows and injections. The real power flow from bus r to bus s with a series impedance z_{rs} between r and s and a shunt admittance at bus r of y_r is

$$P_{rs} = V_r^2 [|Y_{rs}| \cos(\beta_{rs}) + |y_r| \cos(\alpha_r)] - V_r V_s \cos(\theta_r - \theta_s - \beta_{rs})$$

where α_r and β_{rs} are the angles of the admittance of the series and shunt branch respectively:

$$\frac{1}{z_{rs}} = |Y_{rs}| e^{-j\beta_{rs}} \quad y_r = |y_r| e^{j\alpha_r}$$

In general, the measurements depend on the states which are complex voltages $\mathbf{V}$ at angle $\boldsymbol{\theta}$.

$$\mathbf{z} = \mathbf{g}(\mathbf{V}, \boldsymbol{\theta}) + \boldsymbol{\epsilon}$$

Usually, the measurement errors are assumed to have zero mean and are independent, with a diagonal covariance matrix $\mathbf{W}$:

$$E\{\boldsymbol{\epsilon}\} = 0, \quad E\{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\} = \mathbf{W}, \quad w_{ij} = 0, \quad w_{ii} = \sigma_i^2$$

The estimate is formed by minimizing the weighted differences between the actual measurements and the

computed measurement as a function of $\mathbf{V}$ and $\boldsymbol{\theta}$. The weights are the inverse of the covariances giving more weight to better measurements and less to low-quality measurements:

$$J(\mathbf{V}, \boldsymbol{\theta}) = [\mathbf{z} - \mathbf{g}(\mathbf{V}, \boldsymbol{\theta})]^T \mathbf{W}^{-1} [\mathbf{z} - \mathbf{g}(\mathbf{V}, \boldsymbol{\theta})] \\ = \sum_{i=1}^m \frac{(z_i - g_i(V, \theta))^2}{\sigma_i^2}$$

Writing a Taylor expansion about the last iterate $\mathbf{V}^k, \boldsymbol{\theta}^k$:

$$\mathbf{g}(\mathbf{V}, \boldsymbol{\theta}) = \mathbf{g}(\mathbf{V}^k, \boldsymbol{\theta}^k) + \mathbf{G} \begin{bmatrix} \mathbf{V} - \mathbf{V}^k \\ \boldsymbol{\theta} - \boldsymbol{\theta}^k \end{bmatrix}$$

where $\mathbf{G}$ is a matrix of first partial derivatives of the elements of $\mathbf{g}$ with respect to the components of $\mathbf{x}$ evaluated at $\mathbf{x}^k$. If

$$\begin{bmatrix} \Delta \mathbf{V} \\ \Delta \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{V} - \mathbf{V}^k \\ \boldsymbol{\theta} - \boldsymbol{\theta}^k \end{bmatrix}, \quad \Delta \mathbf{z} = \mathbf{z} - \mathbf{g}(\mathbf{V}^k, \boldsymbol{\theta}^k)$$

then one step in the iteration is given by

$$\mathbf{G}^T \mathbf{W}^{-1} \mathbf{G} \begin{bmatrix} \Delta \mathbf{V} \\ \Delta \boldsymbol{\theta} \end{bmatrix} = \mathbf{G}^T \mathbf{W}^{-1} \Delta \mathbf{z} \\ \mathbf{H} \begin{bmatrix} \Delta \mathbf{V} \\ \Delta \boldsymbol{\theta} \end{bmatrix} = \mathbf{G}^T \mathbf{W}^{-1} \Delta \mathbf{z}$$

The gain matrix $\mathbf{H}$ is large and sparse, and the equation is usually solved with Gaussian elimination. The matrix $\mathbf{G}$ is sparse much like the load flow Jacobian. By organizing the measurements into active and reactive power, the equivalent of the fast decoupled load flow can be used to simplify the expression further. With real or active power measurements (sub A) first followed by reactive power (sub R):

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_A \\ \mathbf{z}_R \end{bmatrix}, \quad \mathbf{z}_A = \begin{bmatrix} \mathbf{P}_{km} \\ \mathbf{P}_k \end{bmatrix}, \quad \mathbf{z}_R = \begin{bmatrix} \mathbf{Q}_{km} \\ \mathbf{Q}_k \\ \mathbf{V}_k \end{bmatrix}$$

and states as angles followed by voltage magnitudes to form

$$\begin{bmatrix} \mathbf{H}_{AA} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{RR} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{\theta} \\ \Delta \mathbf{v} \end{bmatrix} = \mathbf{G}^T \mathbf{W}^{-1} \begin{bmatrix} \Delta \mathbf{z}_A \\ \Delta \mathbf{z}_R \end{bmatrix}$$

Using the assumptions from the fast decoupled load flow (small angle difference, voltage magnitudes near 1, and large X/R for the lines), the off diagonal blocks of $\mathbf{H}$ are zero. An approximate $\mathbf{H}$ can even be computed with

angles set equal to zero and voltage magnitudes set equal to one at the cost of more iterations.

The addition of even a few direct measurements of angle to the previous formulation has a number of advantages and creates symmetry in the problem statement. When angle measurements are included, $\mathbf{z}_A$ and $\mathbf{z}_R$ both have three entries [2, 3]:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_A \\ \mathbf{z}_R \end{bmatrix}, \quad \mathbf{z}_A = \begin{bmatrix} \mathbf{P}_{km} \\ \mathbf{P}_k \\ \boldsymbol{\theta}_k \end{bmatrix}, \quad \mathbf{z}_R = \begin{bmatrix} \mathbf{Q}_{km} \\ \mathbf{Q}_k \\ \mathbf{V}_k \end{bmatrix}$$

The matrix $\mathbf{G}$ is modified in an obvious manner, but otherwise, the previous development applies. In order to use the PMU measurements in the conventional SCADA-based system, it is also necessary to put those measurements on the same reference as the convention estimate which has a reference bus. This generally requires the addition of a PMU at the reference bus.

PMUs can also measure the currents in lines connected to the bus. The addition of line currents measurements can be used to compute the bus voltages on adjacent buses if the line model is known. The PMU measurements of both voltage and current have time tags and are different in nature from the SCADA measurements but can improve the conventional estimates even if the time tag is not used and the PMU data is simply treated as additional SCADA data.

PMU-Only Estimators

A modest number of PMUs will permit a PMU-only measurement system for selected parts of the system. For example, a PMU-only estimator for the 500-kV portion of a utility is quite possible. Phasor measurements of the bus voltages and line currents at some 500-kV buses can form the basis of a linear estimate for the 500-kV system that can be updated as often as once a cycle. There have been studies of the number of PMUs required, their location, the phasing in of such systems, and their reliability [4–6]. While the results of these papers are interesting, the actual sequence of installation of PMUs and the locations of the PMUs are frequently far from optimal. Many other issues influence these decisions in the real world.

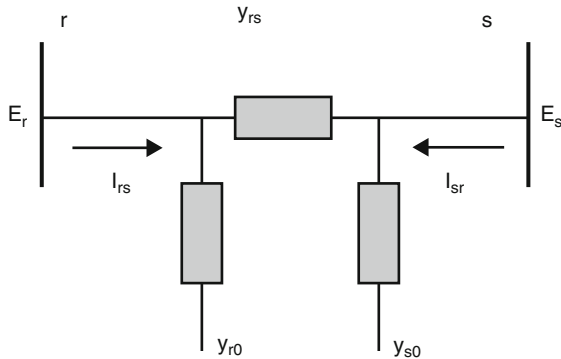
Even recognizing the practical side of the problem, the theoretical results are that fewer PMU are required

than might have been imagined. If we assume that any PMU can measure the bus voltage and all of the line currents at that bus (for a large number of lines, this may not be true for some PMUs, and more than one PMU must be installed in a substation), then the problem of the minimum number of substations required to measure the state of the system can be solved. The most direct solution technique uses the node-node incidence matrix for the system. The matrix A has only 0s and 1s with $A_{ii} = 1$ and $A_{ij} = 1$ if there is a line from bus i to bus j . Using a vector x which has a one in position j if substation j is equipped with PMUs, then for any x the product $y = Ax$ can be formed. Then, y_k is the number of times the voltage at bus k is seen by the various PMUs. If $x = [1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0 \ 1 \ 0]$ indicating PMUs at bus 1, 4, 7, 10, and 13 in the IEEE 14 bus system with A matrix below, then $y = [1 \ 2 \ 1 \ 2 \ 2 \ 1 \ 2 \ 1 \ 3 \ 1 \ 1 \ 1 \ 1]$, indicating that monitoring these five buses can completely observe the system. Nine of the buses are seen by only one PMU, however.

Assuming that only some PMU measurements of bus voltages and line currents are used in the estimate, consider the pi equivalent for a transmission line in Fig. 6. The state of the system is the set of complex bus voltages. The measured currents are linear function of the state, assuming the line parameters are known.

The measurements at both ends of the line are

$$\begin{bmatrix} E_p \\ E_q \\ I_{pq} \\ I_{qp} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ y_{pq} + y_{p0} & -y_{pq} \\ -y_{pq} & y_{pq} + y_{q0} \end{bmatrix} \begin{bmatrix} E_p \\ E_q \end{bmatrix}$$



Wide Area Monitoring, Protection and Control. Figure 6
Pi equivalent of a transmission line

For a more complex network, the measurements can be written as

$$z = \begin{bmatrix} \mathbf{II} \\ \mathbf{YA} + \mathbf{Y}_s \end{bmatrix} [E_b] = \mathbf{BE}_B$$

where $\mathbf{II}$ is related to a unit matrix with diagonal entries equal to one where a voltages is measured and zero otherwise. The second set of measurements are the currents expressed as flows in the shunt elements through y_s and in the series branches. A is an incidence matrix, and y is a matrix of series impedances. The matrices Y and Y_s are complex while the entries in $\mathbf{II}$ and A are only zeros and 1. With $Y = G + jB$ and $Y_s = G_s + jB_s$, both the measurements and states can be written as real and imaginary, and the real products are (real*real-imag*imag) and the imaginary product as (real*imag + imag*real) to form

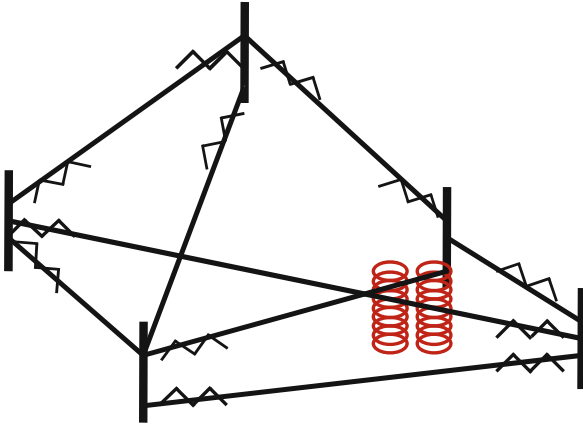
$$\begin{bmatrix} z_r \\ z_x \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \mathbf{II} \\ \mathbf{GA} + \mathbf{G}_s \end{bmatrix} \\ \begin{bmatrix} 0 \\ \mathbf{BA} + \mathbf{B}_s \end{bmatrix} \end{bmatrix} \begin{bmatrix} \begin{bmatrix} 0 \\ -\mathbf{BA} - \mathbf{B}_s \end{bmatrix} \\ \begin{bmatrix} \mathbf{II} \\ \mathbf{GA} + \mathbf{G}_s \end{bmatrix} \end{bmatrix} \begin{bmatrix} E_r \\ E_x \end{bmatrix} + \epsilon$$

$$\hat{x} = (\mathbf{B}^T \mathbf{W}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{W}^{-1} \mathbf{z} = \mathbf{Mz}$$

The estimate is then given by the least squares solution where W is the covariance of the measurement errors. Note that the matrix M is constant, can be computed off-line, and only changes when there is a change in topology (a breaker opening). The resulting estimate is free of the static assumptions and truly dynamic.

Instrument Transformer Calibration for All-PMU Estimators

The network in Fig. 7 shows CTs and CVTs. A precision PT is shown in red. All CTs and CVTs are assumed to have a ratio correction factor: k_i . The true voltages and currents satisfy network equation: $E_t = ZI_t$. The actual measurements are related to true values by the ratio correction factors: $I_m = k_i I_t$, and $E_m = k_e E_t$. A set of E_m and I_m is obtained over several points during daily load variation. With sufficient number of such measurements, a WLS solution for the ratio correction factors can be formed [7]. Large numbers of simulations have shown that the k 's can be determined with high accuracy (equal to that of the good PT).



Wide Area Monitoring, Protection and Control. Figure 7
Three-phase CTs and CVTs with a precision PT

With the notation of the PMU-only estimator, the calibration problem can be written as

$$\mathbf{z} = \mathbf{K}\mathbf{B}\mathbf{x}$$

where $\mathbf{K}$ is a diagonal matrix of ratio correction factors, and $\mathbf{x}$ is the set of complex bus voltages. Both the $\mathbf{k}$'s and the $\mathbf{x}$'s are to be estimated. It is clear that with a $\mathbf{k}$ for every measurement that there are too many unknowns in that if $\mathbf{K}$ and $\mathbf{x}$ are solutions the $\alpha\mathbf{K}$ and $\mathbf{x}/\alpha$ are also solutions. For this reason, one ratio correction factor is taken as 1 (the voltage measurement is taken as being correct); the $\mathbf{k}$ is removed from the set of unknowns. A precision PT or a new high-quality CVT can be used for this purpose.

The estimation of all the $\mathbf{k}$'s and all the $\mathbf{x}$'s from one set of measurements is not possible. It is also necessary for the states to be sufficiently different to estimate all the ratio correction factors. A possible approach is to perform the calibration place over a 12-h period with measurements at a 1-h interval used for the calibration. The calibration is a batch solution for all the states ($12 \times$ the dimension of the state) and the ratio correction factors. It is also assumed that:

1. The transformer ratio correction factors are constant over the simulation period.
2. At least one PT is perfect, which can be used as reference to calibrate other current and voltage transformers.
3. The system model (network impedance matrix) is accurate.

4. The 12 cases are used to get 12 different operating conditions which are sufficiently redundant to estimate the ratio correction factors of transformers.

Each PTs and CTs have a ratio error of the form:

$$\text{Measurement} = M^* \text{True}$$

where M is the ratio correction factor $M = |M| \angle \theta$, where both M and θ are randomly distributed with zero mean. If the precision PT is not perfect but has a known accuracy η , then all the CTs and PTs will have essentially that accuracy.

Seams

Creating a single state estimator for the connection of two large independent systems each with large state estimators is a challenging task. The individual estimators may each involve tens of thousands of buses and represent many years of effort. The estimators may be of different types, come from different vendors, involve different rates, etc. The proposal that two 30,000 bus estimators be merged into one 60,000 estimator is liable to fall on deaf ears. At the simplest level, the problem can be solved by recognizing that the only problem is that the estimators have different references and place two PMUs to measure the difference in the references. A PMU at the reference bus for system 1 and a PMU at the reference bus for system 2 would be sufficient.

A slightly more elaborate solution is to include some buses near the boundary in both systems. This could be done by considering tie lines to be in both systems as shown in Fig. 8.

If ϕ is the angle between the references for the two systems and there are nb boundary buses, a more refined estimate of V can be obtained by averaging over the boundary buses as

$$\hat{\varphi} = \frac{1}{nb} \sum_i^{nb} (\hat{\theta}_{1i} - \hat{\theta}_{2i})$$

where the terms in the sum are the difference between the estimated angles at bus i in the two systems. A further refinement, [8], can be made by forming a performance index including the differences between the measured and calculated real and reactive flows in the tie lines that can be minimized by choice of the angle difference ϕ .

Protection

Several areas of application of synchrophasors to protection are “adaptive” (sections “[Adaptive Security/Dependability](#),” “[Adaptive Out-of-Step](#),” and “[Adaptive Loss-of-Field](#)”). Traditionally, relays were set, and the setting could only be changed by a visit to the substation. Microprocessor relays allowed the relay engineer to change setting remotely, and the concept of adaptive protection introduced the possibility of automatic response to system conditions. “Adaptive protection is a protection philosophy which permits and seeks to make adjustments automatically in various protection functions in order to make them more attuned to prevailing system conditions” [9]. PMUs have made adaptive protection possible. There is an extensive literature on a variety of adaptive protection applications [9–15].

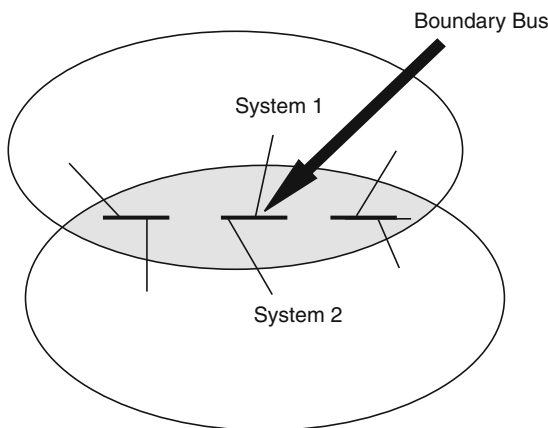
Adaptive Security/Dependability

Traditionally, relays have been biased toward dependability where dependability is defined as the “degree of certainty that a relay or relay system will operate correctly.” There is an inevitable trade-off between dependability and security, “the degree of certainty that a relay or relay system will not operate incorrectly,” that is, incorrect relay operations are either failure to trip in situations in which they should trip or trips in situations when they should not trip. With multiple primary relays (as many as three) and multiple backup

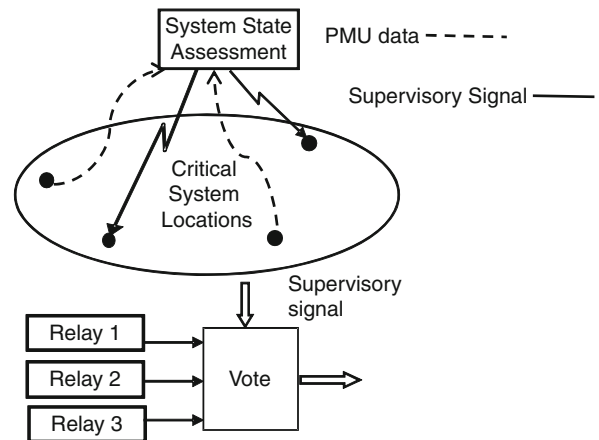
relays (both zone 2 and zone 3), the existing system is heavily biased toward dependability. Examination of NERC data indicates that incorrect relay operations are involved in two third of major disturbances, not necessarily as the initiating event but as a contributing factor to cascading outages. The mechanism of “hidden failures” in relays underlies relay involvement in major disturbances. A hidden failure is “a defect or error in relay that does not manifest itself immediately but which can cause a miss-operation when the system is stressed.” The largest cause of hidden failures is thought to be maintenance. The Internet and Chemical plants also have maintenance-related hidden failures.

The solution to at least some relay involvement in cascading outages is to adjust the security/dependability balance when the system is stressed. One simple way to accomplish this is to require that two of the three relays see the fault before tripping the breaker. This voting procedure only requires altering the way the outputs of the primary relays are combined. The relay logic itself is left unchanged. The PMU contribution to process is to determine if the system is in a stressed condition. [Figure 9](#) shows a voting scheme.

In a specific application, the location of the relays that will vote has to be identified. Important paths or tie lines whose loss would have a serious impact are obvious choices. Then, the location of the phasor measurement and the logic of translating measurements into a decision to vote have to be determined. One successful approach to these two problems is the use



Wide Area Monitoring, Protection and Control. **Figure 8**
Boundary buses between two systems



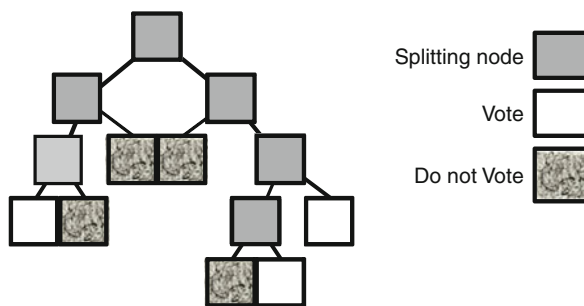
Wide Area Monitoring, Protection and Control. **Figure 9**
An adaptive security/dependability protection scheme

of decision trees. The tree is trained with a selection of a large number of scenarios where it can be determined that a vote would be beneficial combined with a similar set of scenarios in which voting is not necessary. For example, if the path included parallel lines, the loss of an additional line due to a hidden failure can have serious consequences. Many transient stability simulations have to be performed to determine conditions where voting is indicated. The tree is selected with data mining software from an array which has a row for each event and a column for every measurement that might possibly be used. All bus voltages and line currents in the area could be included in the matrix, for example. The last column has a 1 for voting and a 0 for not voting. The data mining software selects the measurements to be used to form a tree.

A tree is shown in Fig. 10. The grey boxes are splitting nodes while the white and textured boxes are terminal nodes. The first splitting node selects one measurement and tests that measurement value against a threshold. The remaining splitting nodes also use only one variable, and it is possible that the same variable can be reused in other splitting nodes with different γ 's. The tree finds both the measurement m_k and the threshold γ . Typically, the initial splitting nodes for these power system problems are quite effective, and the splitting nodes lower in the tree are making fine corrections: $m_k \leq \gamma$ or $m_k > \gamma$ splitting logic.

Monitoring Approach of Apparent Impedances Toward Relay Characteristics

Relay settings are made at the time of commissioning of the substation and the protection system. Once made,



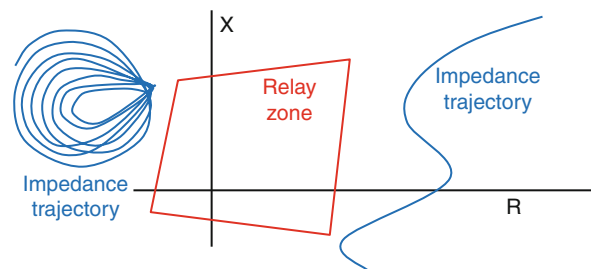
Wide Area Monitoring, Protection and Control.
Figure 10

A decision tree with five splitting nodes

these settings are generally not subjected to periodical review. Although this is not a universal practice, more and more power companies have limited manpower available for such tasks, and it is usually the case that the relay settings are seldom if ever revisited.

It is a fact that many of the relay settings are based upon certain assumptions made about the power system configuration and load levels. This is particularly true about the slower protection functions such as backup zones of distance relays, loss-of-field relays, etc. As years go by, the relay settings which depend upon system conditions may no longer be appropriate for the prevailing power system conditions because of the natural evolution of the power system. As the relay settings become inappropriate, the possibility of false trip of such relays becomes likely. One of the watershed blackout events is the 1965 blackout in North Eastern United States. The settings of some relays involved in this event were made 9 years before the event and had become inappropriate for the system conditions existing in 1965. Misoperation of these relays was a main contributing factor of this blackout.

PMUs employed for WAMS can be programmed to track the behavior of apparent impedance continuously and provide an alarm to the relay engineers that the apparent impedance trajectory recently (on recorded date and time) came within a prescribed distance of the trip characteristic. This is particularly effective when distance relays are used for the protection function. The concept is illustrated in Fig. 11.



Wide Area Monitoring, Protection and Control.

Figure 11

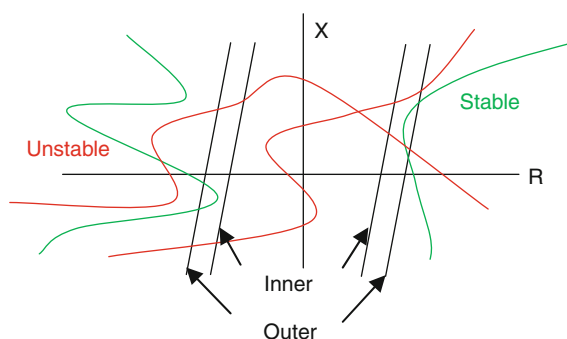
Zone setting of the relay in question is shown in red.

Impedance trajectories are shown in blue, and two types of trajectories – one corresponding to stability swing and the other to load changes – are shown

The figure shows that load changes as well as stability swings did indeed come close to the relay zone, and a message to this effect is sent to the relay engineer responsible for this setting. After reviewing the incident, the relay engineer may decide to review and modify the relay setting, or ignore the alarm confirming that the setting is in fact appropriate for the prevailing power system state.

Adaptive Out-of-Step

Out-of-step (OOS) relaying is used to determine if a major power system disturbance is likely to result in loss of synchronization between groups of generators in the system and lead to island formation of portions of the power system. The present practice is to use distance relays at the terminals of generators or on transmission lines which lie at the electrical center of expected system disturbances. The traditional OOS has a characteristic as shown in Fig. 12 with an inner zone and an outer zone in the relay impedance plane. Using large number of transient simulations performed on the power system, the settings of these relay zones are determined. The inner zone is such that no stable power swing crosses into it, while unstable swings are expected to cross the outer and inner zones in sequence. A timer is set somewhat greater than the



Wide Area Monitoring, Protection and Control.

Figure 12

Traditional out-of-step relay characteristic. All stable swings remain outside the inner zone while the unstable swings cross the outer and the inner zones in time which is shorter than the timer setting of the relay. Stable swings are shown in green, while unstable swings are shown in red

largest dwell time of the unstable impedance trajectory between the two zones. The zone boundaries and the timer are then used to detect whether or not a prevailing disturbance will lead to instability.

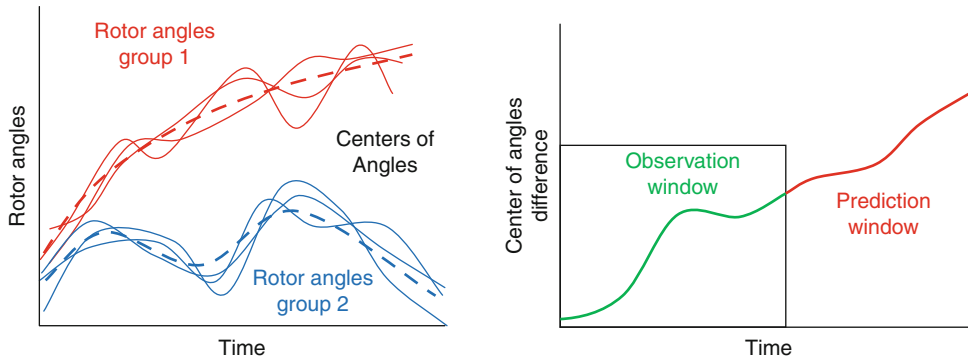
The traditional out-of-step relays are often inadequate when the power system changes or when unanticipated outages occur. In such cases, the relay settings may not be appropriate to the prevailing system conditions, and the OOS relay is likely to misoperate.

Adaptive out-of-step relaying principle proposes to use real-time measurements to determine whether or not an evolving disturbance is going to lead to instability. The most promising real-time measurements to achieve this goal are the phasor measurements provided by the PMUs. In this approach, the first step is to detect the formation of coherent groups of generators and, once the coherent groups have been determined, to predict the likely outcome of power swings between the coherent groups. In order to determine coherency between generators, it is necessary to track the rotor angles of generators in time as the disturbance evolves. The rotor angles are determined from phasor measurements performed at the terminals of the generators and then using the generator equivalent circuit to determine rotor angles. In many cases, it is sufficient to track the behavior of the positive sequence voltage angle at the generator terminals. Figure 13 illustrates the concept of this type of adaptive out-of-step relaying.

It is of course possible that there are some unusual disturbances whereby it is not possible to identify the coherent groups or make a successful prediction based upon an observation window. However, a success rate of around 90% seems to be possible in most power systems and should be considered to be a good outcome of the adaptive out-of-step relaying principle.

Supervision of Backup Zones

Distance relays are used to protect transmission lines. It is common practice to use three zones of distance relays to cover the complete transmission line (with zones 1 and 2) and provide a backup protection for neighboring transmission lines with zone 3. The three zones are given different operating times – zone 1 being the fastest, zone 2 having an operating time of about 300 ms, and zone 3 operating in about 1 s, thus giving

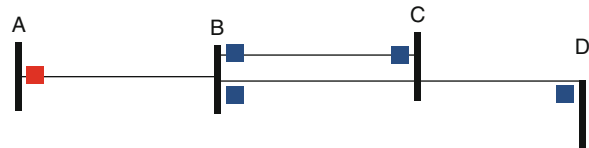


Wide Area Monitoring, Protection and Control. Figure 13

Adaptive out-of-step relaying principle. Generator rotor angles are monitored to determine whether coherent groups are being formed. Once the coherent groups are formed, a prediction regarding the outcome of the swing can be made

enough time for relays of the neighboring lines to operate if there is a fault on those lines. Backup zones of protective relays – in particular the third zone of distance relays – are known to be encroached upon by load swings during system disturbances. As load excursions in the relay protective zones lead to incorrect trips, this usually leads to a worsening situation and accelerates the approach to cascading system failures. Some engineers have recommended getting rid of these overreaching zones completely in order to avoid false trips of this type. However, there are documented cases [16] where it is found to be essential to have zone-3 protections in order to back up neighboring circuits when their primary protection fails for some reason [16]. It then becomes necessary to consider ways in which the overreaching zones of distance relays can be made secure against false trips due to load excursions.

Consider a portion of a power system shown in Fig. 14. Zone 3 of the distance relay at station A (shown in red) is assumed to be picked up because of load excursion during a power system contingency. The decision to be made is whether the zone-3 pickup is due to a fault on lines BC or BD or is in fact caused by a load excursion. If it is a fault on lines BC or BD, at least one of the PMUs located at stations B, C, or D (shown in blue) will indicate a zone-1 fault. If none of the PMUs confirm the existence of such a fault, then zone 3 of the relay at A must be blocked. As zone-3 operating times are of the order of 1 s, there is ample time to get the information from stations B, C, and D to the relay at station A. It should be noted that if an



Wide Area Monitoring, Protection and Control. Figure 14

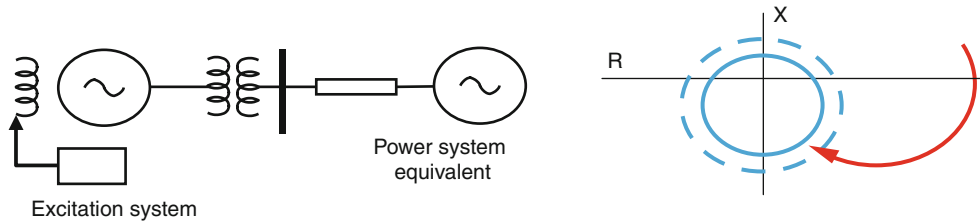
Supervising zone 3 of the distance relay at station A. It is assumed that PMUs exist at stations B, C, and D

Supervising zone 3 of the distance relay at station A. It is assumed that PMUs exist at stations B, C, and D

unbalanced condition is detected at station A, an unbalanced zone-3 fault is indicated, and the blocking of its operation is not appropriate. As load excursions are balanced phenomena, the blocking logic based upon inputs from B, C, or D is appropriate only under balanced network flow conditions.

Adaptive Loss-of-Field

Large generators in power systems are generally protected against operating with reduced field current. The problem faced by these generators is instability when the field current is suddenly reduced due to some problem in the excitation circuit. The traditional protection against loss-of-field (LOF) uses offset impedance relays connected at the terminals of the generator. The relay settings are determined by the impedance of the generator and the equivalent impedance (Thévenin impedance) of the power system to which the generator is connected. A typical loss-of-field characteristic is shown in Fig. 15.



Wide Area Monitoring, Protection and Control. Figure 15

Generator with its excitation system connected to power system. Characteristic of the LOF relay connected to generator terminal is shown on the right

The LOF relay setting consists of two concentric offset impedance circles as shown in the figure in blue color. The inner circle is the actual stability limit, whereas the outer circle shown by dotted line provides an alarm to the operator to take some corrective action before stability limit is reached. The center and radius of the circles are determined by the machine impedance (including the step-up transformer impedance) and the Thévenin impedance of the power system equivalent. A typical impedance trajectory indicating a loss-of-field condition is shown in red: the start of the trajectory is in the first quadrant where the pre-disturbance load is assumed to exist, and as the field current decreases, the impedance trajectory moves in the fourth quadrant and may enter one of the zones of the relay.

During a major disturbance on the power system as transmission lines are lost due to faults or other contingencies, the power system Thévenin impedance may undergo significant changes. These changes should be taken into account and the relay characteristics modified accordingly. If this is not done, it is likely that instability of the generator could result even though the impedance trajectory has not entered the zones set according to “system normal” conditions. An adaptive LOF relay would track the behavior of the power system as it undergoes changes during system disturbances and correct the relay settings automatically to reflect the prevailing state of the Thévenin impedance. In general, the network changes remote from the generator do not affect the Thévenin impedance significantly, and there may be just a few network configurations for which the relay characteristic needs to be changed. The possible contingencies and their corresponding relay settings could be pre-stored as various setting groups in the LOF relay, and when one of these contingencies

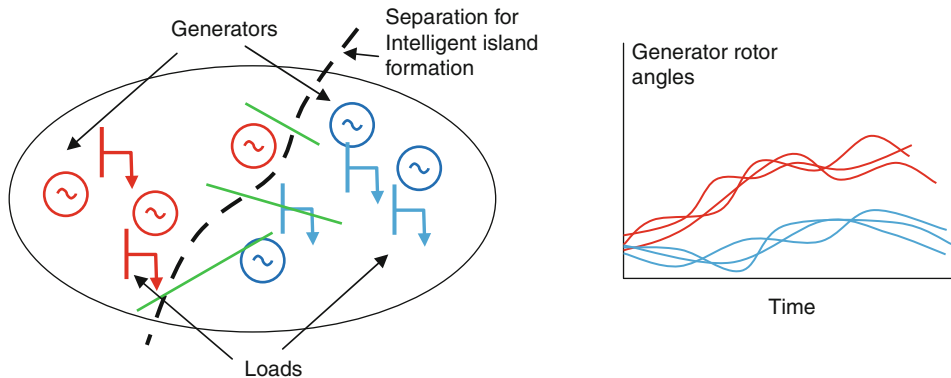
actually occurs on the system, the relay setting changes to the appropriate group.

Intelligent Islanding

When subjected to catastrophic events, generators in the power system may lose synchronism with each other so that they can no longer continue to operate as components of a single power system. Usually, generators tend to maintain synchronism within individual groups, which could get separated from the rest of the power system. Several islands may be formed in this fashion, each island stabilizing as a separate power system with its own generation, transmission system, and load. If the formation of islands is allowed to happen in an uncontrolled fashion, it is possible that unsustainable islands would be formed, and some islands or the entire power system may be subjected to a blackout. This is the motivation for implementing an intelligent islanding scheme.

The first step in such a scheme is to identify groups of generators which remain coherent as a precursor to forming an island. This concept is illustrated in Fig. 16.

Consider the power system in Fig. 16, which has been subjected to a significant contingency, and as a result, groups of generators shown in red and blue develop rotor movements as illustrated in the swing curves shown on the right. Based upon this movement of rotors, it can be concluded that the group of blue generators is a coherent group, and the red generators form another coherent group. The next step is to match the capacity of blue generators with loads shown in blue, thus making up the structure of the proposed island. The red loads are then teamed with red generators, which are also approximately matched in generated power and load. Having identified load and



Wide Area Monitoring, Protection and Control. Figure 16

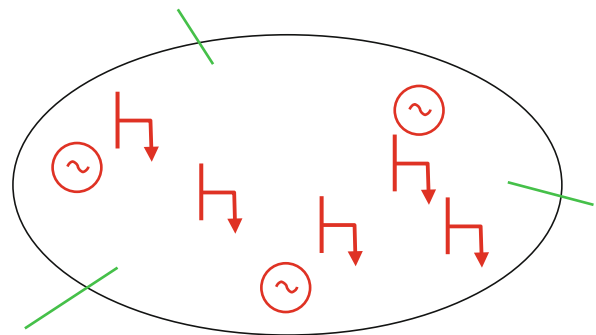
Intelligent islanding principle. Coherency of groups of generators is established, and then possible island with those generators is identified

generation match, the next step is to identify the least number of lines (shown in green) which when opened will create two islands which have the potential to survive as two independent power systems, which can be re-synchronized into a single power system at a later stage.

Intelligent Load Shedding

Most modern power systems are equipped with an underfrequency load shedding scheme by which certain amount of load is shed in order to match available generation with load it can serve. This may become necessary when a block of generation is lost due to faults or other causes, or when the power system has broken up into islands with unequal generation and load. In this scheme, the load shedding is accomplished only after the frequency of the system changes because of the load-generation mismatch. A more intelligent approach to load shedding would be to anticipate the possibility of a load-generation mismatch developing, which may (if unchecked) lead to system islanding. Such an approach requires that an estimate of the load-generation mismatch be formed as soon as it takes place and then shed an appropriate amount of load without waiting for the system to break up into islands and the island frequency to change.

Consider a generic power system shown in Fig. 17. It is a part of an interconnected power system, and it has several tie lines (shown in green) connecting it to the rest of the grid. As the system undergoes a



Wide Area Monitoring, Protection and Control. Figure 17

Intelligent load shedding principle. Changes in tie line flows are used to determine load shedding requirements to balance load and generation within the system

change – such as a loss of some generation – a potential for load-generation imbalance within the boundary of the power system occurs. If changes in the tie line flows are monitored by the Wide Area Monitoring System, it reflects the contributions made by the rest of the grid to aid the power system in riding through this disturbance. From the changes in the tie line flows, an estimate of the impending load-generation imbalance in the system can be made. If it is determined that the potential imbalance is significant, and that if not corrected immediately may lead to islanding of the power system, load shedding could be initiated by supervisory control without waiting for the frequency to decay.

A number of simulation studies need to be performed on the complete grid with various scenarios of generation loss. For the cases which lead to islanding in the simulation studies, different amounts of loads to be shed to prevent the formation of island can be established. The results of these studies can then be used as an aid in determining the load shedding required in actual cases of generation loss.

Power System Control

Prior to synchrophasor measurements, essentially all control in the interconnected power system was based on local measurements. Control that responded to the power system frequency was a possible exception since frequency is a global quantity viewed on a long time scale. Many techniques were developed to deal with the limitation of local measurements. A variety of equivalents, coherent groups, and reduced order models were used to develop control strategies. The existence of wide area measurements with time tags opened up the possibility of control based on some remote measurements. Full state feedback as employed in smaller systems is still impossible, but feedback of important variables can be considered. The choice of the control strategy and placement of remote measurements are problem dependent and are current research topics. Concerns with latency in the signals, the performance of data concentrators, and the design of communication networks have delayed control applications more than some other areas.

Control of Sustained Oscillations

Very low frequency lightly damped low-frequency oscillation has been observed in large power systems for decades. The interactions between groups of machines displaced by large distances tend to be at frequencies less than 1 Hz. The analysis of such modes is based on small signal linear models found by linearizing about an operating point. Such inter-area oscillations are tolerated when they have sufficient damping but can become a serious threat when the damping starts decreasing. The evolution of a 0.27-Hz mode in the WECC disturbance in 1996 is representative. The oscillation began as 0.27 Hz with 7% damping. The tripping of a significant line altered the mode to 0.254 Hz with 3.46% damping which over the

course of about 200 s became 0.252 Hz with 1.2% damping. Shortly thereafter, another line tripped, a generator was dropped, and the mode became unstable [17]. In China, in 2008, two system-wide low-frequency-oscillation events occurred respectively in the South China and Central China grids [18]. A 0.4-Hz oscillation was a precursor to the 2003 Eastern China disturbance [18, 19]. The conventional approach to damping such oscillations is to employ a power system stabilizer (PSS). PSSs use local information such as accelerating power, the generator rotor speed deviation, or frequency deviation as input to the regulator to damp oscillations.

Modern digital power system stabilizers [20] still use local measurements of power and frequency and are aimed at single inter-area oscillations. There have also been studies including remote measurements in a coordinated multi-input environment [21, 22]. The number of remote measurements considered in these studies has been small compared to the number of such measurements that are possible. In rapidly growing systems such as the Chinese power system, the limitation of PSSs that are designed for specific modes is more apparent. The need to retune PSSs has prompted investigation of techniques to adapt PSS operation to changing system conditions [23, 24]. The most ambitious approach is to coordinate PSSs, FACTS devices, and DC lines in stabilizing inter-area oscillations [25–28]. Linear feedback based on various robust control techniques have been employed to deal with power system nonlinearity and uncertainty. Time delays have been modeled as uncertain parameters [25], unmeasured portion of the system was represented by unmodeled dynamics [27], and LMI techniques used to design robust H_2/H_∞ controllers [28–33].

Control of Large Oscillations

Faults that remain on the system longer than a “critical fault clearing time” can cause oscillations large enough that the nonlinear nature of the system is exposed. The oscillations are described by a coupled set of nonlinear differential equations. Depending on the complexity of the representation of the generators, each machine can be described from 2 to 20 or more state variables. At the simplest level, the “classical” machine model involves the rotor angle and speed (the derivative of the angle).

If the rotor angle is denoted by δ , the “swing” equation for a single machine in a per-unit system takes the form:

$$M\ddot{\delta} + D\dot{\delta} = P_m - P_e$$

where M is the inertia of the machine, D is the damping, P_m is the mechanical input power, and P_e is the electrical output power. The equation is nonlinear since the electrical power is a function of angle differences, i.e., P_e depends trigonometrically on δ . In equilibrium, $P_m = P_e$, and δ is constant. The equal area criterion in section “[Adaptive Out-of-Step](#)” is developed from examination of the swing equation for a single machine. Even in using reduced order modeling in large systems, there are hundreds of coupled nonlinear differential equations to be solved in order to study the transient stability problem. The equations are coupled by the P_e term because every machine connected to machine n has an electrical power contribution to P_e . If the fault duration is too long, the system is moved away from equilibrium on the “fault on trajectory” and begins to oscillate or swings when the fault is cleared. In large systems, the time for one period of a transient swing is a second or more, suggesting that time is available for PMU measurements to be used in control of these oscillations. The means available for control are somewhat limited. Fast valving of steam turbines; insertion of a dynamic brake; generator, reactor, and line switching; control of DC line flow; and FACTS devices are possible. Specific applications of PMU inputs to some of these controls have been reported [34–40]. The use of PMU measurements in these control schemes should improve their performance.

In one study, the control involved the DC lines and was discrete in nature. The PMU measurements were mapped into regions where specific discrete control actions were taken [41]. Many off-line simulations were used to generate the mapping, but no time was involved in real-time computing a control action.

Remedial Action Schemes

Remedial Action Schemes (also known as Special Protection Schemes or System Integrity Protection Schemes – SIPS) are designed to address a problem caused by a relaying action. For example, the loss of one of three parallel lines could overload the remaining

two lines. The scheme might not always be enabled but could be armed based on system conditions. If the flow in the lines was large enough, the control scheme could drop load and/or generation when a relaying action removed any of the parallel lines. In some cases, operators arm the RAS scheme. All SIPS are essentially specially designed control and protection systems and can be enhanced with PMU input [42]. The addition of PMU information for SIPS could be managed using the decision tree approach used for adaptive security/dependability in section “[Adaptive Security/Dependability](#)” The proper coordination of large numbers of these schemes is an even more challenging problem.

System Restoration

Synch check relays are used to ensure that two AC systems are synchronized. For a system to be synchronized, the frequency, phase angle, and voltage must be within preset limits. Wide area measurements could be used to perform these same functions across breakers used to close tie lines between islands reclosing when the angle differences are within acceptable bounds. If the angles exceed acceptable limits, a generation-load rescheduling within the islands could bring the quantities within limits. The large number of unsuccessful attempts to reclose in Europe in November 2006 gives testimony to the need for real-time angle measurements in restoration [33]. Restoration could only be achieved after the phase angles of separated areas become acceptable.

Artificial Neural Networks (ANN) have been proposed to aid in system restoration [45]. The ANNs can be trained to determine island boundaries for restoration and to determine a sequence of switching operations which would lead to restoration, taking into account overvoltages due to light load conditions and problem caused by cold-load pickup.

Future Directions

Wide Area Measurement Systems utilizing Phasor Measurement Units have become the power system sensors of choice. These systems deliver precise simultaneous measurements of key power system variables from which reliable real-time models of power systems can be created and intelligent monitoring, protection, and control decisions could be implemented.

These measurements also offer one of the most reliable tools for postmortem analyses of catastrophic events taking place on the power systems. From such analyses follows an understanding of realistic models of power system elements and more realistic simulation tools.

The future of this technology holds the promise that power systems of the future will be made more secure – that is, they will be less immune to catastrophic failures – and in rare cases when such failures do occur, their impact will be considerably reduced in its physical extent and duration. This will be made possible through protection and control tasks which take into account precise knowledge of the current state of the system and take actions which are best suited to the prevailing power system conditions.

It is already clear that the technology is making an impact on all major power producers throughout the world. As the world brings into the mainstream of electric power establishment greater number of alternative energy sources, dependence on this technology is going to be even greater as engineers face the challenge of integrating these newer-generation systems which behave in unpredictable and intermittent fashion. Application of this technology in coming years will usher in an era of cleaner, abundant, and reliable supply of electricity to future generations.

Bibliography

Primary Literature

- Dopazo JF et al (1975) Implementation of the AEP real-time monitoring system. *IEEE Trans Power Apparatus Syst* 95:1618–1629
- Thorp JS, Phadke AG, Karimi KJ (1985) Real-time voltage phasor measurements for static state estimation. *IEEE Trans Power Apparatus Syst* 104:3098–3107
- Phadke AG, Thorp JS, Karimi KJ (1986) State estimation with phasor measurements. *IEEE Trans Power Syst* 1:233–241
- Nuki RF, Phadke AG (2005) Phasor measurement placement techniques for complete and incomplete observability. *IEEE Trans Power Deliv* 20:2381–2388
- Gou B, Abur A (2001) An improved measurement placement algorithm for network observability. *IEEE Trans Power Syst* 16:819–824
- Abur A (2005) Optimal placement of phasor measurements units for state estimation. PSERC Publication 06–58, Oct 2005
- Zhou M (2008) Phasor measurement unit calibration and applications in state estimation. Ph.D. dissertation, Virginia Tech
- Jeffers R (2007) Wide area state estimation techniques using phasor measurement data. Virginia Tech Report prepared for Tennessee Valley Authority
- Centeno V, De La Ree J, Phadke AG, Michel G, Murphy RJ, Burnett R (1993) Adaptive out-of-step relaying using phasor measurement techniques. *IEEE Comput Appl Power* 6:2–17
- Horowitz SH, Phadke AG, Thorp JS (1988) Adaptive transmission system relaying. *IEEE Trans Power Deliv* 3:1436–1445
- Rockefeller GD, Wagner CL, Linders JR, Hicks KL, Rizi DT (1988) Adaptive transmission relaying concepts for improved performance. *IEEE Trans Power Deliv* 3:1446–1458
- Hedding RA, Mekic F (2007) Advanced multi-terminal line current differential relaying and applications. In: *Protective Relay Engineers*, 60th annual conference, College Station, 27–29 Mar 2007, pp 102–109
- Lien K-P, Liu C-W, Jiang J-A, Chen C, Yu C-S (2005) A novel fault location algorithm for multi-terminal lines using phasor measurement units. In: *Proceedings of the 37th annual North American power symposium*, Ames, Iowa, 23–25 Oct 2005, pp 576–581
- Lin Y-H, Liu C-W, Jiang J-A (2000) An adaptive fault locator for transmission lines tapped with a source of generation-using synchronized voltage and current phasors, pp 1379–1383
- Yu C-S, Liu C-W, Yu S-L, Jiang J-A (2002) A novel adaptive PMU-based transmission-line relay-design and EMTP simulation results. *IEEE Trans Power Deliv* 17:33–46
- Horowitz SH, Phadke AG (2006) Third zone revisited. *IEEE Trans Power Deliv* 21:23–29
- Hauer AJF, Trudnowski DJ, DeSteese JG (2007) A perspective on WAMS analysis tools for tracking of oscillatory dynamics. In: *IEEE Power Engineering General meeting*, Tampa, 24–28 June 2007, pp 1–10
- Dong M, Yang D, Huang Y (2008) Analysis on the low frequency oscillation incidents measured by WAMS of central China power grid during the 2008 ice hazard. *Cen China Electr Power* 21:22–25
- Zhu F, Liu ZH, Chu L (2004) Achievement and experience of improving power system stability by PSS/excitation control in China. In: *IEEE Power Engineering Society general meeting*, Denver, 10 June 2004, pp 1767–1771
- Kamwa I, Grondin R, Trudel G (2005) IEEE PSS 2B versus PSS4B: the limits of performance of modern power system stabilizers. *IEEE Trans Power Syst* 20:903–915
- Kamwa I, Gerin-Lajoie L (2000) State-space system identification-towards MIMO models for modal analysis and optimization of bulk power systems. *IEEE Trans Power Syst* 15:326–335
- Zhang Y, Bose A (2008) Design of wide-area damping controls for interarea oscillations. *IEEE Trans Power Syst* 23:1136–1143
- Yu GL, Zhang H, Xie H, Wang CG (2007) Wide-area measurement-based robust control of power system considering signal's delay and incompleteness. In: *Power Engineering Society general meeting*, Tampa, 24–28 June 2007
- Hu G, He R, Yang H, Wang P, Ma R (2005) Iterative Prony method based power system low frequency oscillation mode

- analysis and PSS design. In: Transmission and distribution conference and exhibition: Asia and Pacific, Dalian, 2005, pp 1–6
25. Gibbard MJ, Vowles DJ, Pourbeik P (2000) Interactions between, and effectiveness of, power system stabilizers, and FACTS device stabilizers in multimachine systems. *IEEE Trans Power Syst* 15:748–755
 26. Yu GI, Zhang BH, Xie H, Wang GC (2007) Wide-area measurement-based nonlinear robust control of power system considering signal's delay and incompleteness. In: Power Engineering Society general meeting, Tampa, 24–28 June 2007
 27. Li X, Yan Q, Hu Y (2008) Decentralized coordinated robust adaptive control for AC/DC interconnected power system based on WAMS. *IEEE Asia-Pacific Conf Circuits Syst* 2008: 1896–1899
 28. Dou C, Jin S, Bo Z, Jia Q, Xue H (2007) Decentralized stabilizer design for model-based uncertain large interconnected power systems based on WAMS. In: IPES, Singapore, 3–6 Dec 2007, pp 799–806
 29. Rao PS, Sen I (2000) Robust pole placement stabilizer design using linear matrix inequalities. *IEEE Trans Power Syst* 15:313–319
 30. Soliman M, Emara H, Elshafei A, Bahga, A, Malik OP (2008) Robust output feedback power system stabilizer design: an LMI approach. In: IEEE Power and Energy Society general meeting conversion and delivery of electrical energy in the 21st century, Pittsburgh, 20–24 Jul 2008, pp 1–8
 31. De Campos VAF, Da Cruz JJ, Zanetta LC (2006) Pole placement and robust adjustment of power systems stabilizers through linear matrix inequalities. In: Power systems conference and exposition, Atlanta, 29 Oct–1 Nov 2006, pp 2180–2187
 32. Bensenouci A, Abdel Ghany AM (2007) Mixed H_∞/H_2 with pole placement design of robust LMI-based put feedback controllers for multi-area load frequency control. In: EUROCON, Warsaw, 9–12 Sept 2007, pp 1561–1566
 33. Jabr RA, Pal BC, Martin N (2010) A sequential conic programming approach for the coordinated and robust design of power system stabilizers. *IEEE Trans Power Syst* 25:1627–1636
 34. Rostamkolai N (1986) Adaptive optimal control of AC/DC systems. Ph.D. dissertation, Virginia Tech
 35. Manansala EC (1989) Adaptive power system control. Ph.D. dissertation, Virginia Tech, Aug 1989
 36. Rostamkolai N, Phadke AG, Thorp JS, Long WF (1988) Measurement based optimal control of high voltage AC/DC systems. *IEEE Trans Power Syst* 3:1139–1145
 37. Manansala EC, Phadke AG (1991) An optimal centralized controller with nonlinear voltage control. *Electr Mach Power Syst* 19:139–156
 38. Mili L, Baldwin T, Phadke AG (1991) Phasor measurements for voltage and transient stability monitoring and control. Workshop on application of advanced mathematics to power systems, San Francisco, 4–6 Sept 1991
 39. Snyder AF et al (1988) Inter-area oscillation damping with power system stabilizers and synchronized phasor measurements. In: PowerCon 1988, China
 40. Smith MA (1994) Improved dynamic stability using FACTS devices with phasor measurement feedback. MS thesis, Virginia Tech
 41. Rovnyak S, Taylor CW, Thorp JS (1995) Real-time transient stability prediction—possibilities for on-line automatic database generation and classifier training. In: 2nd IFAC symposium on control of power plants and power systems, Cancun
 42. Madani V, Taylor E, Erwin D, Meklin A, Adamik M (2007) High speed control scheme to prevent instability of a large multi-unit power plant. In: Protective Relay Engineers, 2007: 60th annual conference, Clemson power system conference, pp 271–282
 43. Sattinger W (2005) Developments in UCTE and Switzerland WAMC course at ETH Zürich, 30 Aug–1 Sept 2005
 44. ENTSOE (2003) European network of transmission system operators for electricity. Final Report of the Investigation Committee, 28 Sept 2003 Blackout in Italy. <https://www.entsoe.eu/index.php?id=59>
 45. Bretas AS, Phadke AG (2003) Artificial neural networks in power system restoration. *IEEE Trans Power Deliv* 18: 1181–1186

Books and Reviews

- Horowitz SH, Phadke AG (2008) Power system relaying, 3rd edn. Wiley, Chichester/Hoboken
- IEEE Standard C37.118-2005 (Revision of IEEE Std 1344-1995) IEEE standard for synchrophasors for power systems
- Kundur P (1994) Power system stability and control. McGraw-Hill, New York
- Phadke AG, Thorp JS (2008) Synchronized phasor measurements and their applications. Springer, New York
- Phadke AG, Thorp JS (2009) Computer relaying for power systems, 2nd edn. Wiley, Chichester
- Rogers G (2000) Power system oscillations. Kluwer, Norwell

Wind Power Balancing

LENNART SÖDER¹, HANNELE HOLTINEN²

¹Royal Institute of Technology, Electric Power Systems, Stockholm, Sweden

²Technical Research Centre of Finland VTT, VTT, Finland

Article Outline

Glossary

Definition of the Subject

Introduction

Physical Balancing
 Electricity Markets and Production Planning
 How Wind Power Changes the Balancing of Systems
 Results from Integration Studies
 Conclusions and Future Directions
 Bibliography

Glossary

Primary control The instant reaction in a power system to keep the continuous balance between production and consumption.

Secondary control Normally automatic changes in the production system in order to release used primary controlled units.

Control area A change in production or consumption in a certain control area will cause an automatic response from the secondary controlled units in the same control area.

Regulating capacity The power plants and consumers which can change their production or consumption when requested in order to keep the balance.

Balancing market A market where regulating capacity can be bid in and the best offers are then used when balancing is needed.

Balancing reserves Regulating capacity that is bid into a balancing market.

Balancing area A change in production or consumption in a certain balancing area will cause an automatic or manual response from the regulating capacity in the same balancing area.

Definition of the Subject

Wind power balancing considers the challenge to keep a continuous balance between total production and total consumption in a power system. The varying wind speed introduces special considerations concerning how to keep this balance in an efficient way. This article is mainly focused on wind power as a part of a larger power system.

Introduction

In a power system there is always a challenge to keep the balance between total production and total consumption. This challenge has existed ever since the first power systems were installed more than 100 years ago.

Since 1970 the amount of wind power has grown from a marginal source to levels of more than 50% during some situations in certain parts of some systems. This increase is expected to continue, which means new challenges and requirement of new methods to keep the requested balance in an efficient way.

This chapter provides an overview about the challenges that are faced with the integration of large-scale wind power into the power system. There are always changes in production and consumption in all power systems since the load changes continuously and there are sometimes sudden outages in power plants and/or power lines. Despite these changes, the continuous balance between production and consumption is always kept. With larger amounts of wind power, the net changes in the controllable part of the production and consumption system will increase since wind power changes will increase the overall changes seen by the rest of the system. The challenges include balancing of changes and keeping enough margins for reacting to changes depending on uncertain forecasts. Also, methods for estimation of the size of the increased challenges need to be addressed, as well as efficient market rules for an economic operation.

This chapter first describes the challenge of how to keep the physical balance between production and consumption in any power system, [Physical Balancing](#). This is followed by a section concerning how the planning of the balancing can be organized in an electricity market, [Electricity Markets and Production Planning](#). The next section includes a description of how a power system is affected when the share of wind power increases. Finally results from several large-scale integration studies and conclusions from the chapter are presented, “[Results from Integration Studies](#)” to “[Conclusions and Future Directions](#).”

Physical Balancing

The starting point is that electricity cannot be stored. Today, some electricity storage technologies exist but are still not cost effective to be used in large power systems. This means that as soon as electricity is consumed it has to be generated and vice versa, that is, as soon as electricity is generated it has to be consumed. This applies always and is a physical law that cannot be changed. In practice, this means that as soon as a lamp

is switched on, the corresponding electricity has to be produced somewhere else. The other way around it is true that – when a wind or coal power station (or any other power station) generates power – at exactly the same time, the electricity has to be consumed somewhere else. This means electricity cannot “disappear”! There are of course always losses in the power system but these are usually only a few percent (in, e.g., Sweden ca. 10%) of the consumption if the entire chain from the generation feed-in via power lines all the way to the consumer is considered. Losses are caused by power lines and transformers that get warmer when a current flows through them. In practice, this means that total production (= what generators feed into the power system) is always larger than what consumers take out of the power system. Sometime, the words “power shortage” or “excess power” are used in this context. These terms, however, rather refer to economic issues. From a physical perspective, there is no “excess power” or “power shortage.” “Power shortage” means that somebody *would like* to consume electricity, but that there is no power available at the consumer’s location or for the price the consumer is willing to pay. As opposed to this, “excess power” refers to a situation where there are power stations that *could* produce power if consumers were willing to consume more and to pay the price for this power.

It is naturally a challenge for the power system to maintain an exact balance between production and total consumption including power system losses. The biggest challenge for a power system is often a situation where – for some reason – there is a sudden unplanned stop in a large power station. In that case a production capacity of, for example, about 1,000 MW of power generation can be lost, which corresponds to the amount of power generated in a larger nuclear power station. In the exact moment when such an incident occurs, the design rule is normally that all consumers continue to be supplied with power as usually, that is, as the failed power station cannot supply them any longer, the electricity has to come from another power source. A power source that is always available as “reserve” power are all the spinning parts of the generators, generators that are on-line producing electricity (in hydro-power stations, nuclear power stations, coal and gas power stations, and others) of the entire power system, such as, for example, the whole

Nordic power system, the Central European EU UCTE System etc. The spinning parts include the rotor, the shaft, and the turbine. At the very moment a large power station comes to a sudden stop, the power is supplied from the energy stored in all these spinning parts, and then reserves start to operate in different time scales. A bicycle can be used to illustrate how the balance in a power system is maintained. This is shown in Fig. 1.

Regulating Capacity and Reserves in Power Systems

In power systems it is, as discussed above, crucial to maintain both the flexibility and reserves necessary to keep the system operating under a range of conditions. These conditions include possible unscheduled plant outages, as well as predictable and uncertain variations in load and in primary generation resources including wind. Day-ahead scheduling for unit commitment is done according to the load (and wind) forecast and real-time load following is accomplished with operational reserves.

Continuous Balancing: Load Following Capacity

Since production always has to equal consumption a power system needs to be *flexible*. As consumption varies, even production has to vary accordingly. Here, power stations that are available for continuously following power consumption are defined as *load following capacity*. This group comprises a large number of power stations as the difference in consumption between day and night can amount to several thousand MW. *Power consumption* refers here to the net consumption that has to be supplied by power stations whose production is controlled specifically with that goal. In a first step, wind power is considered as “negative power consumption,” that is, the other power stations have to supply the net consumption = difference between actual power consumption and wind power production.

Figure 2 shows a continuous curve that describes the power consumption that has to be continuously supplied where the *load changes* are met by controllable power production, that is, *load following capacity*. Most of the load following in the figure is scheduled day ahead. In the figure the hydro power is used for daily regulation, while nuclear power during this week is

used for weekly regulation (to help to follow the lower consumption during weekend).

The figure shows an example of the continuous variations of net consumption. It always exactly equals production that, therefore, also has to vary. There are some phenomena to consider:

- When net consumption is at its lowest level it can only increase. This is due to that consumption increases and/or wind power decreases. At this level, also controllable production is at a low level, resulting in that there are several power stations with margins to increase their production. In practice, it is not possible to say when the lowest production level is reached; therefore, it is necessary to have standby capacity to decrease production even then.
- Challenge for following an increasing power consumption. There are often some power plants on-line that can increase their production level and this can be done at short notice. In addition it is possible to start up power plants that are not on-line. For many power plants this can take 4–12 h. The challenge is to keep enough power plants on-line to be able to follow the changes needed in short notice. Some power systems may have limitations as to how fast production can be increased. This is a particularly big challenge in systems with thermal power plants, for example, coal-fired power stations, as these need time to heat up and may have limitations regarding the ramping speed in MW/min. Hydro power often can move from zero to full production in a short time, 5–10 min, but how short this time is depends on a number of different factors, such as water supply, conditions due to water legislation, current operational mode and technical solutions of the individual plant. It is also possible to use demand response – lowering consumption in these cases.
- Challenge for following a decreasing power consumption. Technically it is possible to *decrease* production of the power plants on-line to the desired level in most cases, but this can in some cases result in energy losses when, for instance, steam has to be dumped or water has to be spilled. Most power plants have a minimum operation limit before they have to be shut down, and shutting

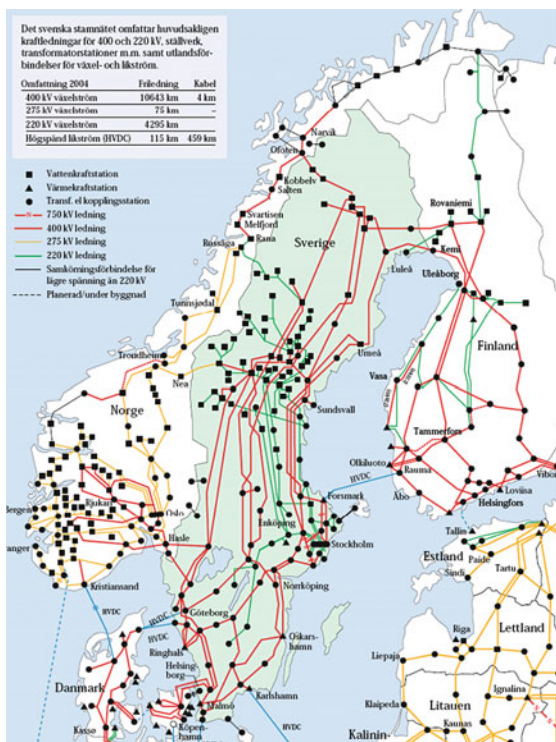
down can be a lengthy process. It is also possible to use demand response – increasing consumption in these cases.

Continuous Balancing: Operational Reserves

Day-ahead scheduling can cover most of the load following. However, one challenge when operating a power system is that there are no perfect forecasts. A lot of things may occur, for instance, power line interruptions, unpredicted changes in consumption and unpredicted changes in wind power or other generation. Independent of what happens, controllable power stations have to change production in that case because total production has to equal total consumption at all times. This means, there have to be reserves, that is, margins to change production level in case of such events. The assumption is to have reserves in power stations, but it is also possible to have reserves in the form of changed consumption and/or changed transmission on controllable power lines to power systems in neighboring countries. However, if you change the transmission on controllable lines you move the balancing requirement to the other side of the line and the production (or consumption) has to be changed there instead.

On the bicycle in Fig. 1, another cyclist has to pedal if somebody stops pedaling. The cyclist that now pedals more was not able to pedal to the maximum before. The *margin* this cyclist has can thus be called a *reserve* that this person has in case somebody else stops pedaling or somebody else breaks stronger. The same occurs in a power system. If a power station stops working, if wind power plants decrease production, or if somebody increases power consumption, there has to be some other power station that can increase production. If a nuclear power station with a usual production of 1,000 MW suddenly stops, other power stations have to have margins to directly increase production by 1,000 MW. If wind power production decreases by 200 MW, there have to be other power stations that take over that production to supply power consumption if at the same time there are no changes in consumption as production always is equal to consumption.

It is important, though, to take into account, that “reserve” means that a power station stands by to produce more. On the bicycle, there is one person

Bicycle	Power system															
	Det svenska stamnet omfattar huvudsakligen kraftledningar för 400 och 220 kV, ställverk, transformatorstationer m.m. samt utlandsförbindelser för växel- och likström. <table><tr><th>Utförande 2004</th><th>Färdledning</th><th>Kabel</th></tr><tr><td>400 kV växelström</td><td>10643 km</td><td>4 km</td></tr><tr><td>220 kV växelström</td><td>75 km</td><td>-</td></tr><tr><td>220 kV växelström</td><td>4295 km</td><td>-</td></tr><tr><td>Högspänning likström (HVDC)</td><td>115 km</td><td>459 km</td></tr></table> ■ Vattenkraftstation ▲ Värmekraftstation ● Transf. elkopplingsstation — 750 kV ledning — 400 kV ledning — 220 kV ledning — 220 kV ledning — Samkörningsförbindelse för lägre spänning än 220 kV - - - Planerad/under byggnad 	Utförande 2004	Färdledning	Kabel	400 kV växelström	10643 km	4 km	220 kV växelström	75 km	-	220 kV växelström	4295 km	-	Högspänning likström (HVDC)	115 km	459 km
Utförande 2004	Färdledning	Kabel														
400 kV växelström	10643 km	4 km														
220 kV växelström	75 km	-														
220 kV växelström	4295 km	-														
Högspänning likström (HVDC)	115 km	459 km														
It is assumed that some people on the bicycle pedal and others break. Air resistance and wheel resistance are “losses” and are part of the breaking process. If cyclists pedal the whole time exactly as much as they break, the bicycle will move at a constant speed.	This corresponds to a situation where total production equals total consumption in the power system (e.g., the Scandinavian power system shown in this figure). This balance means that all large generators (with so-called synchronous generators where the electrical frequency is directly related to how fast the generator spins) spin with the same frequency and you get a constant frequency in the entire power system. It is normally 50 Hz, that is, the alternating voltage changes polarity between plus and minus exactly 50 times per second.															
If somebody stops pedalling because his leg suddenly hurts, the bike slows down. It does not directly stop because it has a certain inertia due to the weight of the people on the bike and of the bike itself.	This corresponds to a situation where a power station suddenly stops production. The power system starts to slow down and the frequency slightly decreases. It starts decreasing in the same way the bike starts slowing down. However, the frequency does not directly drop as there is a large amount of kinetic energy in the turbines, etc. that drive the generators.															
When the bike’s speed (the same speed across the entire bicycle) decreases, there are some cyclists who pedal and have a speedometer. If they notice that the speed goes down they can pedal stronger in order to compensate for the cyclist whose leg hurts. The remaining cyclists increase their power on the pedal to a level that keeps the speed constant. Again there is a balance between pedallers and breakers.	This corresponds to a situation where there are frequency meters installed in several Scandinavian power stations, mainly hydropower stations, that continuously measure the electric frequency (which is the same in the entire Scandinavian power system). When the frequency decreases, the power production in these power stations increase in order to compensate for the reduced production in one of the other power stations. After that the frequency is stable again. All this happens automatically, independent of who has caused the imbalance: whether a power station fails, the wind speed drops in a wind power plant, somebody increases power consumption, or somebody starts-up a power station, etc. After that, suitable power stations are started up by manual regulation in order to restore the automatic reserve.															
In a large power system such as, for example, the Scandinavian one, it is important to take into account limitations in the transmission capacity in order to not overload the lines when variations in one area are compensated with changes in another area. On the bike, the chain can jump off if everybody in the front pedals and everybody else breaks.																

Wind Power Balancing. Figure 1

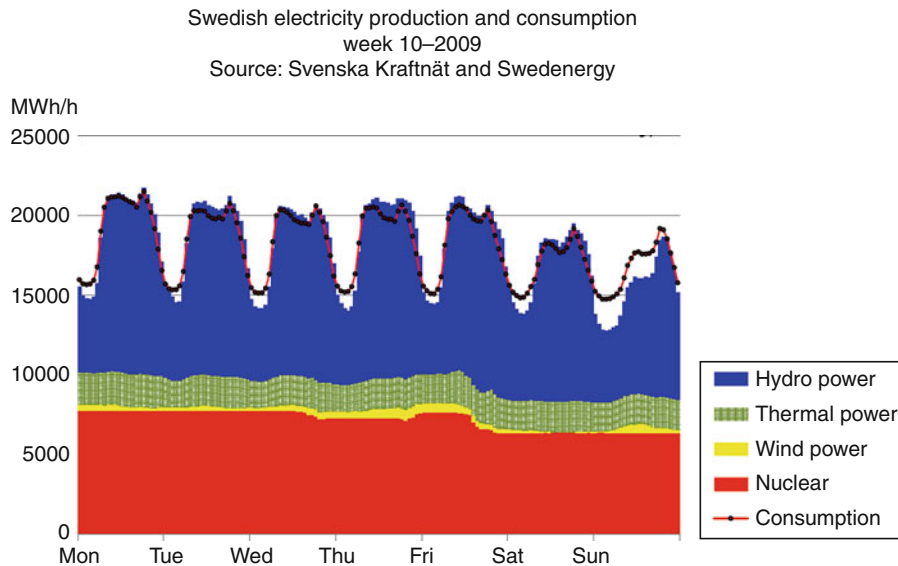
Illustration of how to keep the balance in a power system [1]

that *stands by* to pedal more in case this is needed, that is, no effort is involved in *standing by*. The same occurs in the power system; the power station that maintains *margins=reserves* does not produce any power, but *would be able* to do so if required. Sometime there is a misconception of those power stations that are kept as a reserve increase emissions continuously or use water unnecessarily. However, there might be a certain effect if you operate a power station at a lower efficiency. In order to create fast reserves, coal-fired power stations are sometimes operated at low levels (with lower efficiency), even though production costs may be higher than optimal (the person on the bicycle cannot be at home and sleep). A power station (or part of a power station) that works as a *reserve* cannot produce any power; as the actual goal is *not* to produce any power, but just to *stand by* to do so if required. It is mainly power production that causes, for instance, coal-fired power stations to emit carbon dioxide and water from hydro-power station dams to flow through the turbines. It is also important that there are steady reserves for down-regulation if load decreases or wind power increases. To do this cost efficiently this should be done

without this down-regulation causing unnecessary spill in hydro power, for example.

It is a big challenge to design a system that keeps reserves in a way that is economically efficient. Some power stations (e.g., hydropower stations that are operated at optimum efficiency or gas turbines on hold) hardly incur any costs for *keeping reserves*, but there will be costs *if they are used*. In other power stations (e.g., coal-fired power stations that operate at a low level), there are costs for keeping reserves (energy losses due to less efficiency), but efficiency increases if the reserves are used, that is, the production increases. An efficient approach to handling reserves must therefore take into account the costs of keeping them, the costs of using them as well as an estimate of the probability of being used. An efficient way of handling reserves has to include efficient market signals that give correct incentives.

The examples refer to reserves for production increase, but there are also corresponding examples for reserves regarding the decrease of production. Figure 3 shows an example of a (wrong) forecast at the beginning of a time period for hours 5–24, and an updated forecast for hours 15–24.

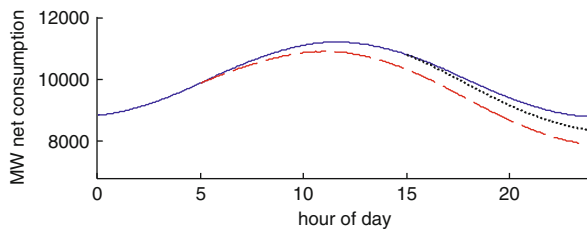


Wind Power Balancing. Figure 2

Power production and consumption in Sweden during week 10, 2009. The difference between production and consumption corresponds to the power traded with neighboring countries. Power consumption changes continuously during the week and has to be supplied by controllable production

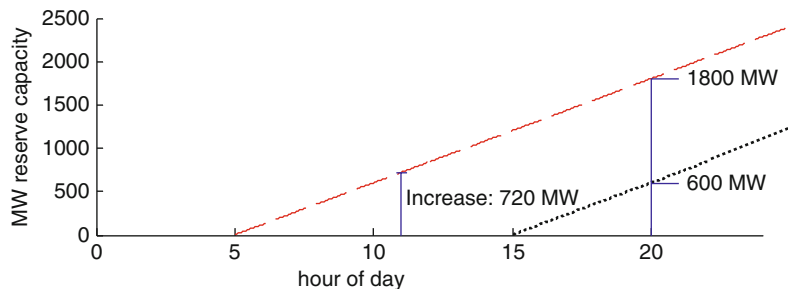
The *reserves* that have to be kept depend on how fast they are available and how large the demand is for different time horizons, that is, what are the dimensioning events. Here the time horizon is of key importance: It may happen, for instance, that it will get very cold (very high power consumption in a system with large amounts of electric heating) within a few days. But if all available resources can be started within 4 h, a decision has to be taken only 4 h before the high power consumption will occur. However, it may be necessary for the respective power station to call in employees and this may require a notice of 24 h.

If you continuously update the forecast, this means that you also can continuously decrease the uncertainty of the forecast for a certain hour. Figure 4 shows the needed production increase in relation to the forecasts in Fig. 3. It shows that for the forecast for the hours 5–24, the forecast error for hour 20 is 1,800 MW; but in the updated forecast for hour 15–24, the forecast error for the same hour has decreased to 600 MW. If the



Wind Power Balancing. Figure 3

Actual power consumption=production in other than wind power plants (*blue solid line*) and forecasts from time $t = 5$ (*red dashed line*) and time $t = 15$ (*black dotted line*)



Wind Power Balancing. Figure 4

Forecast errors from the first forecast for hour 5–24 (*red dashed line*) and for the updated one for hour 15–24 (*black dotted line*). This corresponds to the production increase required within the respective time interval. At hour 5, there have to be reserves of 720 MW that can be started up within 6 h

power stations need a start-up time of 15 h, a reserve capacity of 1,800 MW is required at hour 5; but if 5 h are sufficient for the start-up, the required reserve capacity is only 600 MW based on the forecast error at hour 15.

In the following, the reserves are grouped according to how fast they are needed, which also is connected to the forecast certainty for the same time horizon.

Reserve Categories Reserves can be divided into different categories according to the time scale within which they are operating. There are different practices and terminology in different power systems, but generally speaking there is first an instantaneous reserve that is maintained in power plants operating on-line and is activated automatically to correct for frequency fluctuations. These can be further divided to working at several time scales (from some seconds to a minute). Then there is a somewhat slower reserve that is activated 10–15 min after the occurrence of frequency deviation from nominal frequency. It replaces the instantaneous reserve. This reserve consists mostly of rapidly starting gas turbine power plants, hydro (pump) storage plants, unloaded generation already online, and load shedding can also be used.

There is wide agreement on the type of reserves that must be carried and the time frames that are relevant, but there is no universal agreement on nomenclature. Here the following terminology is adopted:

- Contingency reserve (also called disturbance reserve): capacity that is online and synchronized to guard against unforeseen equipment failure.

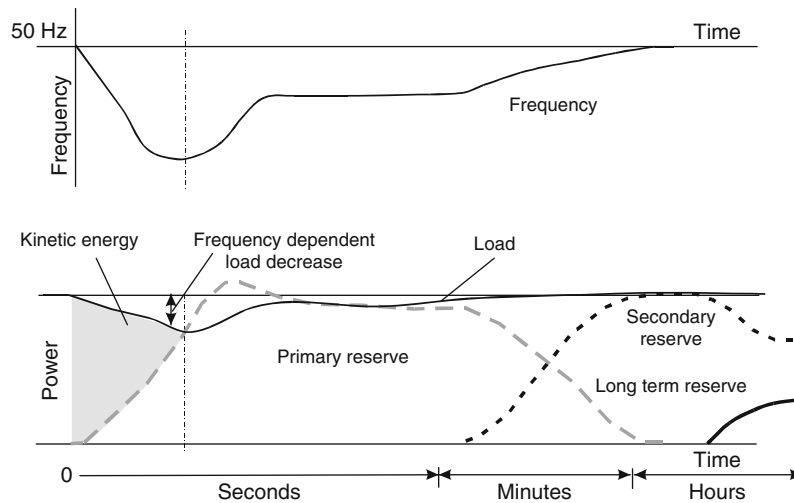
The amount of contingency reserve is typically based on the largest potential source of failure, either the largest generator or tie line in the system. Contingency reserve is held independently of the variability of loads, and is typically shared among participants within the balancing area, reducing the overall contingency requirements for the individual entities. Some portion of the contingency reserve can be non-spinning, but must be capable of responding quickly if needed. This reserve includes both primary (automatic) and secondary/tertiary (manual) reserves.

- Primary/secondary control (also called regulating reserve, response, instantaneous/momentary reserve, primary/secondary reserve): fast fluctuations in system load require fast generation response. Regulation consists of flexible generation that is responding with a speed/droop characteristic to frequency changes (primary control, timescale several seconds) or connected to automatic generation control (AGC) systems trying to adjust the production to certain frequency targets (secondary control, timescale several seconds to minutes). The AGC will send control signals to the participating regulating units as needed to offload used primary reserves so that the aggregate load is balanced with generation. Units that provide primary/secondary reserve are usually run near the mid-point of the operating range. When a regulating unit is required to deviate from that preferred operating point for a longer period of time, adjustments are made to the units on economic dispatch so that the regulating unit can return to its mid-point operating setting.
- Tertiary control (also called load-following reserve tertiary reserve, minute reserve): This consists of generation that may be called on to increase or decrease output over slower time frames than regulating units. The units that are available for load following must be committed in advance, or capable of quickly starting and synchronizing with the grid, typically within 10 min. Depending on the resource mix that is available to the power system operator, load-following reserve can be either spinning or non-spinning, or as is most common, will be a combination of both. Slow-start generation must be committed prior to becoming available for dispatch or to supply reserves.
- Planning reserve (also called long-term reserve): This type of reserve covers not only operating requirements, but recognizes that not all of the installed capacity may be available when needed. Planning reserve is usually found to be in the range of 12–20% of peak load, depending on the system and the planning criteria in use at the locale in question.

An example of how the reserves operate is illustrated in Fig. 5. It shows the frequency of the system and activation of reserves as a function of time when a large power plant is disconnected from the power system. Activation of reserves divides the reserves into primary reserve (also called instantaneous or automatic reserve), secondary reserve (also called fast reserve) and long-term reserve (also called slow or tertiary reserve). Primary/secondary reserve is activated automatically by frequency fluctuations. Tertiary reserve is activated manually or automatically in 10–15 min after the occurrence of frequency deviation from nominal frequency. It replaces the primary reserve and it will be in operation until long-term reserves substitute it as seen from Fig. 5. The tertiary reserve consists of spinning reserve (hydro or thermal plants in part load operation) and standing reserve (rapidly starting gas turbine power plants and load shedding).

It is fairly straightforward to determine contingency reserve required because the largest contingency is a known specific event. Primary/secondary/tertiary control performance and reserve allocation is more difficult to specify. It is not practical and it is not necessary for any single balancing area within an interconnection to continuously perfectly match generation and load. For example, the Control Performance Standard 1 (CPS1) in the USA recognizes this and places an annual statistical limit on the 1-min product of balancing area imbalance and system frequency deviation. Note that this is a performance metric, not reserve requirement. A balancing authority must have “enough” reserves to meet the statistical performance standards, however much that is.

The key aspect of load/frequency control on interconnected systems mainly addresses three time frames and these are generally considered as Primary Control (or governor response, designed to arrest



Wind Power Balancing. Figure 5

Activation of power reserves and frequency of power system as a function of time when a large power plant is disconnected from the power system [2]

frequency decay), Secondary Control (designed to reset the primary control response as well as to balance supply and demand) and Tertiary Control.

Primary Control

The primary control considers the capability of the system to meet sudden load and/or generation changes. The primary control is separate for each synchronous power system. A synchronous power system is a power system where all producers and consumers are connected to each other through transformers and AC transmission and distribution lines. Anything from a diesel generator set supplying a single load to a multinational grid as the Nordel system (which connect Norway, Sweden, Finland and the eastern part of Denmark), or the central European UCTE grid can constitute a synchronous grid. An AC line has to have the same electric frequency at both ends of the line. If there were different frequencies at the ends then the voltage angle shift would increase until it reaches 180° , resulting in unacceptable large currents on the line. The same is valid for transformers. The conclusion is that in a synchronous grid the average electric frequency must be the same. For short periods the frequency can be slightly different, but that will cause large power oscillations in the grid, which can result in serious disturbances.

In a synchronous grid there has to be some kind of primary control. A short description of this type of control, cf. Fig. 1, is given here:

- Assume that a power system is in balance, that is, electricity generation and consumption are equal.
- At a certain time the consumption increases, while the generation remains the same.
- There is rotational energy stored in all synchronous machines connected to the system, more exactly in the rotors and the connected turbine shafts. Since the generation and consumption always have to be in balance the load increase is compensated using the rotational energy, which causes the rotor speed to decrease in the synchronous machines.
- In synchronous machines, there is a strong connection between rotor speed and electric frequency. This means that all synchronous machines in a synchronous grid, in principle, rotate at the same speed. The reduced rotor speed therefore results in a frequency decrease in the grid.
- In some power plants there is frequency sensitive equipment, which sends a signal so the generation increases when the frequency decreases. These power plants are referred to as primary controlled units.
- As long as the frequency keeps falling the primary controlled units will increase their generation.

Finally, the balance between generation and consumption will be restored and the frequency will be stable again. It should be noted that the new stable frequency will be lower than the original frequency!

In the description above it is assumed that it was a load increase that triggered the primary control. A corresponding course of events will come up if there is a failure in a generating unit – if a generating unit is stopped and the load is constant then the remaining units in the system will have to increase their generation. A load decrease and/or a generation increase also causes similar courses of events. The only difference is that the frequency is increasing, which results in generation decreases in some power plants.

Permissible Frequency Range In each power system there is a nominal frequency – North and South America usually 60 Hz, whereas most of the remaining world uses 50 Hz – and it is not possible to allow too large deviations from the nominal frequency, because there is then a risk that important components in the power system are damaged. If, for example, the electric frequency would coincide with the harmonic oscillation frequency of turbine blades or shafts, then these power plants have to be shut down to avoid expensive repairs. Another problem is that generators and transformers are designed for the nominal frequency and a deviating frequency can cause heating of the windings, resulting in damages. Finally, there might be certain kinds of loads that do not work satisfactory if the frequency deviates too much from the nominal.

Speed-Droop Characteristics As explained above there are some power plants which automatically adjust their generation whenever the frequency of the system changes. The speed-droop characteristic is a measure of how large the change in generation is for a certain change in frequency

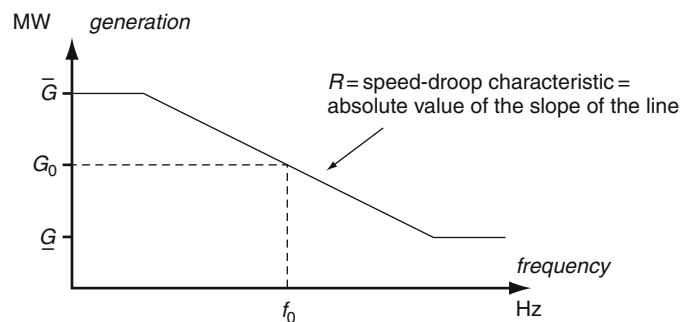
Definition The speed-droop characteristic, R , indicates how the generation in a power plant is changed when the frequency changes. It is measured in MW/Hz. If G_0 is the generation at nominal frequency, $f_0 = 50$ Hz, it holds that $G = G_0 - R(f - f_0)$.

The notion of speed-droop characteristics is illustrated in Fig. 6. In practice it is often inconvenient to use the formula given in the definition. It is easier to just calculate how the changes in generation and frequency relate to each other:

$$\Delta G = R \cdot \Delta f.$$

Then it is simple to conclude whether a certain change corresponds to an increase or a decrease; there must be a shortage of generation if the frequency is falling and therefore the primary control must result in an increase of generation, etc.

When calculating frequency changes, it is necessary to also consider the reserves available in the primary controlled power plants; it is impossible to increase the generation of a power plant operating at maximal capacity regardless of the speed-droop characteristic. This means that when primary control is considered, both the total speed droop and the total available primary reserves are to be defined. The total speed-droop is in, for example, the Nordic



Wind Power Balancing. Figure 6

The relation between frequency and generation in a power plant participating in the primary control

synchronous system divided between the different areas such as Sweden often keeps around 2,500 MW/Hz, Norway 2,000 MW/Hz, Finland 1,250 MW/Hz, and Eastern Denmark 250 MW/Hz at nominal frequency, 50 Hz. This means a total speed droop of 6,000 MW/Hz. This implies, for example, that a sudden outage in Sweden of 600 MW decreases the frequency with 0,1 Hz which causes a compensation of 250 MW in Sweden, 200 MW in Norway, 125 MW in Finland, and 25 MW in Eastern Denmark. This example illustrates the important issue that when many regions or countries are interconnected, then an outage in one part of the system will affect the generation in other parts of the system, which means that margins also have to be kept on the transmission lines between the different parts. The overall aim of the primary control is to balance all sudden changes, so the same process is valid also for changes in, for example, load or wind power production. In some systems the contingency reserve will be activated at a lower frequency than the normal frequency range, to keep the two reserve categories separate.

Secondary and Tertiary Control

The ability of the power system to maintain the balance between production and consumption starts from automatic frequency regulation of the primary regulation and continues with secondary and tertiary control. A consequence of primary regulation is that there still is a frequency deviation, even if the frequency drop or increase has been stopped. The use of primary regulation also means that reserve margins are utilized. A consequence is also that the transmission between different areas will change as shown above. Note that if you have a *constant* frequency – independent of whether it is 50.0 or 49.95 Hz – production and consumption are balanced, similar to the bicycle that moves at constant speed. However, the system is designed to operate within a range of $50 \pm \varepsilon$ Hz (ε is often in the range 0.01–0.5 Hz); and outside these limits, there are smaller margins as part of the margins have already been used. If the frequency approaches the limits, it is desirable to move the system back to the middle of this range. The only way to do this is to start-up new power stations (or disconnect consumption) for frequencies that are below 50.0 Hz or disconnect

power stations for frequencies above 50.0 Hz. This has to be done quite often, several times per hour, as consumption – and subsequently the frequency – varies continuously.

Secondary Control is often under, so-called, automatic generator control, AGC, such that a control block/area that is in imbalance following a disturbance autonomously restores the frequency to the target frequency and the power interchanges with adjacent control blocks to their preset levels. Thus, there is a continuous interaction of the primary and secondary regulation. Secondary control timescale of operation is several, for example, 5–10 min.

The function of a control area is essential for the operation of a power system. The result of the AGC is often that all changes within the control area will within around 5–10 min be automatically balanced within the same control area. If the area contains, for instance, wind power and the wind power decreases with 100 MW when consumption remains constant, then the AGC can control the other power stations in the same area so they automatically increase their production with 100 MW. It is of course technically necessary to increase the power production somewhere to compensate for the decreased wind power production. But if the control areas are very small, then margins have to be kept in each area and the balancing is automatically performed internally in each area although it may be a better economic solution to balance outside the control area. The term *balancing area* can be used to define the areas which have a common balancing. Assume that there are two control areas, but one have an overall (manual) system that studies the economic operation of the system and changes the set value of the transmission between the areas (used for the AGC in each area) if this is to prefer for an economic operation of the two areas. In this example there are then two control areas but only one balancing area.

The Scandinavian Nordel system is not divided into control areas, although the size of the system is about a circle with the diameter of around 1,500 km. In this system there is no automatic AGC, but a system is implemented where all producers provide bids where they offer to increase or decrease their power production within 5–10 min for a certain cost. When an increase of power production is needed (=secondary-tertiary control) then the cheapest bid is accepted that

fulfils the transmission constraints. This means for instance that if wind power decrease in Eastern Denmark is causing an imbalance for the system, then the decrease is first compensated with the primary control, and then the primary control could be replaced with an increase in Finnish hydro power 1,500 km away if they provide the lowest cost bid for their up-regulation and the transmission grid operators find it suitable to change the transmission in this way. In a congested transmission situation a bid which is closer to the imbalance source has to be accepted.

Electricity Markets and Production Planning

A power system always has to have a balance between production, G_i and consumption, D_i , that is, for a power system consisting of i load points in the transmission and distribution systems.

$$\sum_i G_i = \sum_i D_i + \text{losses}$$

This equation is valid independent of the length of the period, that is, if G_i and D_i are measured as MWh/h, MWs/s or MW. The aim of the market is to keep the continuous balance between production and consumption in an efficient and reliable way. First the general structure of operation planning will be presented.

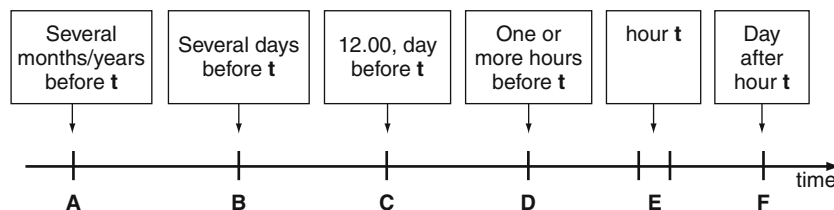
The planning (i.e., before a specific hour) and operation (i.e., within the specific hour) of a power system can be illustrated as in Fig. 7. The overall aim of any market design is to obtain a reliable power supply and an efficient use of available resources.

Failure to keep the electricity system running has serious and costly consequences, thus the reliability of the system has to be kept at a very high level. Security of

supply needs to be maintained in both the short and the long term. This means maintaining both the flexibility and reserves necessary to keep the power system operating under a range of conditions, including peak load situations. These conditions include credible plant outages (disturbance reserves) as well as predictable and uncertain variations in load and in primary generation resources, including wind (operational reserves). In the above presented general process, this means that in all stages A–E enough margins have to be taken into account in order to maintain the security.

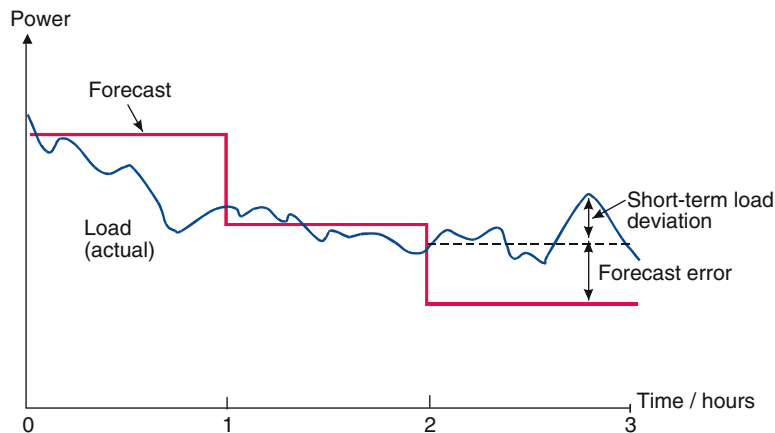
Load following is performed partly beforehand (stage A–D) as scheduling and dispatch of power plants according to the load forecast and partly by operational reserves to balance the load forecast errors. Scheduling includes planning the start-ups and shut downs of slower power plants, called unit commitment, in the time scale of 3–12 h (stage C–D). Optimizing the use of the water stored in hydro power reservoirs, the hydro power plants take into account an even longer scheduling horizon. The scheduling can be based on electricity market operation, where bids for production and bids for consumption based on forecasts are made. Figure 8 shows an example of the actual load in the power system over 3 h compared to hourly forecasted load, denoting forecast errors and short-term load deviations in the system.

In addition to frequency control, the voltage level is managed to prevent under- and over-voltages in the power system and to minimize grid losses. Frequency is a wide area quantity. Measures can be taken anywhere in the system to maintain the balance, as long as transmission capacity is available. Voltage is a local quantity, and voltage management should be taken care of in the vicinity of imbalances. In order to manage the voltage level during disturbances, reactive reserves in power



Wind Power Balancing. Figure 7

Planning and operation of a power system



Wind Power Balancing. Figure 8

Example of actual load in the system over 3 h compared to forecasted load

plants are allocated to the system. These reserves are mainly used as primary reserves in order to guarantee that the voltage level of the power system remains stable during disturbances. Power plants and special equipment, for example, capacitors and reactors, control the reactive power. The voltage ratio of different voltage levels can be adjusted by tap-changers in power transformers. This requires a reactive power flow between different voltage levels.

Seasonal/Yearly Planning

Physical aim: This is especially important for a hydro-based system, where the aim of this planning is to decide how much water (in reservoirs and inflow) should be stored for future use. Too much water stored for months ahead, or next year could lead to future spillage, while too little stored water could lead to lack of energy which could cause water deficit in the future. In a thermal power system, the seasonal planning includes scheduling and coordination of unit maintenance.

Market process: In a hydro power system the tool for optimal storage is the *water value* which is an evaluation of the water for different reservoir levels and different time periods. Different water value in different periods is an indication that it is profitable to change the storage plan. In a liberalized market each hydro-power company makes its own estimation of its water value and the result is used internally for planning and

externally for trading. For thermal unit maintenance the aim is normally to schedule maintenance to low load periods.

Weekly Planning

Physical aim: The aim of this planning in a hydro system with reservoirs is to decide how the available water for the week in the specific region (available from inflow and storage where the use for the storage is planned in stage A) should be distributed between the days of the week. In a thermal power system there may be some planning concerning amount of staff needed for operation of units needed to supply forecasted load, cf. Fig. 2 where the nuclear power is downregulated during the weekend.

Market process: The tool for optimal storage in a hydro system is also here the water value, which is estimated by each actor on the market.

Daily Planning: Day-Ahead Market

Physical aim: The aim of this planning is to decide for each plant how it should be operated, that is, how much power that should be produced in each power plant during each hour in the coming day. In a hydro power system the available water at different price levels for the whole day is a result from the weekly planning. In the planning, the consumption level, production in other sources, and transmission limits have to be considered. In a thermal power system there are

two parts: the *unit commitment*, which includes which units have to be on line for production and/or reserve keeping, and the *economic scheduling*, which includes the planned production level in each unit.

Market process: In this description it is assumed that the day is divided into whole hours. This is the case in, for example, the Nordic power market, but there may also be other time steps as, for example, $\frac{1}{2}$ h. One challenging part is that the power plants have to be operated in an economical way at the same time as the transmission limits have to be considered. This is done slightly different in different systems, depending on if congestion management is handled with implicit auction, explicit auction, counter trading, or another method [3]. The description below is based on implicit auction=market splitting which is used in the Nordic system. There are then four stages here:

- C1. First the transmission limits for each hour and each bottleneck are decided. These limits are calculated by the organization responsible for the transmission system, for example, TSO:s. It is based on an assumed production and consumption level in each area or each node together with a dimensioning rule including the N-1 criterion (= the system has to stand an outage even if the largest unit or the heaviest loaded transmission line fails) and margins considering uncertainties. The margins have to consider that the limits consider MW limits, but in stage C3, these limits are used as hourly limits considering MWh/h. This description is based as an assumption of the so-called *market splitting* method, also denoted *implicit auction*. There are other versions, for example, *market coupling*, but in any system there must be an interaction between transmission operation and margins and plant scheduling. The more close in time the transmission and production scheduling are performed, the more economic efficient the whole system is operated.
- C2. The second stage in a liberalized market with several producers is the production/consumption bidding process where the aim is to put bids to a spot market (Nordpool in the Nordic market) concerning how much power in total could be produced during each hour in each company at which price. The bids concerning each hour have

to be delivered the day before the real trading day. This is the day-ahead market: in Nordic system the bids have to be delivered before noon a certain day for all hours during the coming day.

- C3. The third stage is the calculation of which bids that should be accepted considering transmission limits between bidding areas (Nordic system is based on areas), or limits between nodes where a node-based system (based on DC load flow) is used. This means that as many cheap bids as possible from stage 2 are accepted regarding the transmission limits from stage 1. This calculation is made by the exchange (Nordpool in the Nordic market).
- C4. The fourth stage is the operation planning which is performed individually by the owners of the different production sources after the result of the bids. This means that the accepted amount of production bids for each hour have to be met with produced power in some way (combination of own production in each company and bilateral trade with other companies). This stage includes production planning with the physical aim as above.

Intraday Market

Physical aim: To get an economical operation of all power plants for each specific hour during the day. Here better forecasts for each hour can be used, since the decisions are taken closer to the hour. Especially in a power system with wind power, this phase is important since the 2–3 h wind power forecasts are much better than the 24 h forecasts. Any larger deviations to plans made previous day can be corrected in intraday market.

Market process: It is performed after the spot market for each hour has been closed, that is, when the day-ahead planning is finished. This includes in the Nordic power market the Elbas Market and other bilateral trade. The main aim is to cover forecast errors concerning load levels, wind power production levels, production costs and/or outages for each hour. Bids on the Elbas market are distributed electronically and accepted when they are found economical (is there an English term for this kind of market?). There are the same stages here as concerning the daily market:

- D1. The transmission limits for each bottleneck are set by transmission system operator. This means that they can be updated compared to what is previously set for the daily market.
- D2. Bids concerning production in a certain area can be delivered.
- D3. For each geographical area only bids that can be used in that area (considering transmission limits and prescheduled trading) can be accepted.
- D4. When a bid is accepted, both the seller and buyer have to consider this in their planning.

Hourly Operation

Physical aim: Within each hour the system balance between production and consumption has to be kept continuously in an economical and reliable way, considering production and transmission limits.

Market process: Final schedules are given to the TSOs 30–90 min before the operating hour and the balancing after that is the responsibility of TSOs. This is to avoid cross regulation where each actor regulates their schedules as only the system net imbalances need to be regulated, resulting in a more efficient use of regulation capacity. The continuous balance is kept with primary control (frequency controlled power plants that keep enough margins for this), which is distributed between different power plants in different regions. Used primary reserves are off-loaded using secondary/tertiary reserves, which in the market mean acceptance of bids from the so-called regulating market, or balancing market. These bids can also be used for offloading of bottlenecks (if the amount of power flowing on lines between two areas is too close to physical limits) in order to keep enough margins also on these bottlenecks. There may also be an automatic system (Automatic Generation Control – AGC) which changes the production in some selected units in order to keep the balance and offload units activated by the primary control. The stages in the operation are:

- E1. The transmission limits for each bottleneck are estimated by the transmission system operator. Here the MW limits (not MWh/h) are considered.

- E3. When bids are needed (for total system needs and/or for offloading of bottlenecks) they are accepted by the operator. If an AGC system is active, then some power plants are activated automatically.
- E4. When a bid is accepted the company that put the bid has to activate it directly, or to be more specific, start the procedure of activation. It may take several minutes until the full bid is activated.

Post-trading

Aim: The hour has now passed which means that the physical balance has already been kept, as described for step E above. But what happened in step E was that the system operator, for example, bought power from one producer to supply the system since some consumers that perhaps had a contract with another producer increased their consumption. The aim of the post-trading is to arrange economical balance so it matches the contracts on physical delivery on the market.

Market process: This is mainly relevant in liberalized markets. The contracts for each actor must have been reported to the system operator. All actors connected to the grid must have a contract concerning the economical responsibility for the production and/or consumption. This is denoted *balance responsibility*. If, for example, retailer R1 has a contract to deliver all power to consumer D1 (i.e., retailer R1 is *balance responsible* for consumer D1) then this information has to be reported to the system operator. Then retailer R1 can write contracts concerning production with producer P1 and P2. The aim of steps A–D for retailer R1 is then to try to buy the expected consumption of consumer D1 for hour h from producer P1 and P2 or some other producers. These contracts then have to be reported to the system operator. The basic market tool, the *balance responsibility*, means that there is an *economic incentive* for the market actors to keep an hourly balance between production and consumption. The physical intra-hour balance is kept by the system operators. If, for example, R1 has not bought enough from P1 or P2, then some other producer have delivered the power needed for D1, and the post-trading then implies that R1 has to pay for this extra power. The post-trading, balance settlement, can have several rules. For example, there can be extra penalties for all deviations or then the imbalance pricing can reflect the real

cost for imbalances, as €/MWh, for each hour. Furthermore, even following a price signal from the balancing market, the imbalance pricing can be a one-price system or two-price system. In one-price system there is a cost for those that have been causing the imbalance with a deviation to the same direction up/down as balancing market demand and a revenue for those that have had their deviation to the opposite side. In two-price system, the producers get the spot price during the hours when the deviation has been to the opposite side of system imbalance (balancing market demand) and the balancing price is only used when the deviation is to the same side as balancing market demand.

It can be noted that all the steps above, Planning, A–D, and real-time operation, E, also exist in not liberalized markets, but step F is mainly interesting when there are several actors, which is not the case in many vertically integrated markets.

How Wind Power Changes the Balancing of Systems

The varying production patterns of wind generation changes the scheduling and unit commitment of the other production plants and use of transmission between regions. These changes will be made according to wind power production forecasts 2–40 h ahead and either losses or benefits are introduced to the system, compared with the situation without wind. Most of the energy produced by wind power will impact the scheduling process, but there is a crucial impact on the operating reserves as well. Depending on the prediction accuracy, parts of the wind power production variations remain unpredicted or are predicted incorrectly, and there will also remain subhourly changes in output. This is what has to be handled by load following reserve or the balancing market.

When speaking about the impacts of wind, it is important to know just how variable and predictable the resource is. Wind power production is characterized by variations on all time scales: seconds, minutes, hours, days, months and years. Predictability increases in shorter time scales, but even the short-term variations are to some extent unpredictable. There are means to reduce the variability of wind power production. Staggered starts and stops from full power as well

as reduced (positive) ramp rates can reduce the most extreme fluctuations, in both magnitude and frequency, over short time scales [4]. This reduction is at the expense of production losses, so any frequent use of these options should be weighed against other measures (in other production units) in cost effectiveness.

Power systems usually cover larger areas. It is then relevant to look at the aggregated wind power production from the whole system or balancing area. Large geographical spreading of installed wind power will reduce the variability and increase the predictability of the aggregate wind power production. The variability and predictability can differ in different regions of the world. When assessing the amount of variability it is also important to know how concentrated or distributed the wind power is built in the power system area.

To estimate the impact of wind power on power system balancing, it has to be studied on a power system or balancing area basis. Every change in wind output does not need to be matched one-for-one by a change in another generating unit moving in the opposite direction. Rather, it is the total system aggregation of the variations in all loads and generators that has to be balanced. The relative increase in system fluctuations due to wind power depends on the wind penetration level – what portion of consumption is covered by wind power production. Also, each system is different: the levels of load variability, as well as the flexibility in the system, differ from region to region and from country to country. This means that even for systems with the same wind penetration level the challenge of wind integration and any costs related to that will be different [5, 6].

Also the operational routines of the power system are relevant – how often the forecasts of load and wind are updated, for example. If a re-dispatch based on forecast update is done in 4–6 h, this would lower the costs of integrating wind compared with scheduling based on only day-ahead forecasts. Emerging intraday markets reflect this, giving the opportunity for hourly updates. The costs will depend on the marginal costs for providing regulation or mitigation methods used in the power system as well as on the market rules. The way the power system is operated regarding the time lapse between forecast schedules and delivery will impact the degree of uncertainty wind power will bring about.

There are already several power systems and control areas coping with large amounts of wind power [7]. The experience from Denmark, Spain, Portugal and Ireland that have integrated 9–20% of wind energy (of yearly electricity demand) show that wind power production will only have to be curtailed at some rare instances, and that TSOs need on-line information of both the production and demand levels as well as respective forecasts in their control rooms. Spain and Portugal have launched centers for distributed energy that convey data to TSOs and even can react to control needs [8].

How Variable and Uncertain is Wind Power Production?

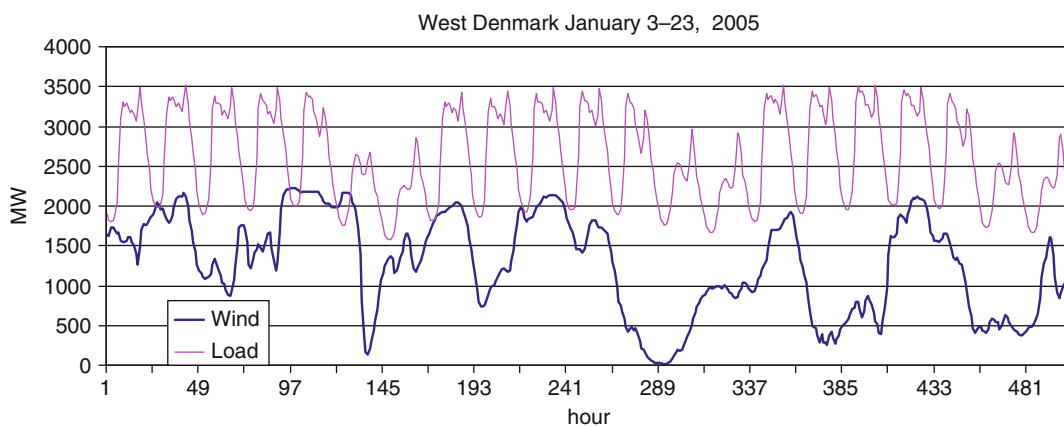
When estimating the system impacts of wind power, it is crucial to describe the variability and uncertainty of wind power in the right way for the aggregated wind power production of the relevant area. Relevant questions are: how much is the variability on average and in extreme situations? How much are the forecast errors for different times ahead, on average and in extreme situation? As there is considerable smoothing effect when combining the output of individual turbines and wind farms, especially in the minute and hourly time scale, it is important to know how large the area in question is and how dispersed or concentrated the wind power capacity will be distributed in the area.

Variability of Distributed Wind Power Production

The variability of wind has been widely studied. Recently also measured large-scale wind power production data has become available to give insight on the variability that is relevant for power system operation (Fig. 9). In-depth information about the variability can be found in [9–15].

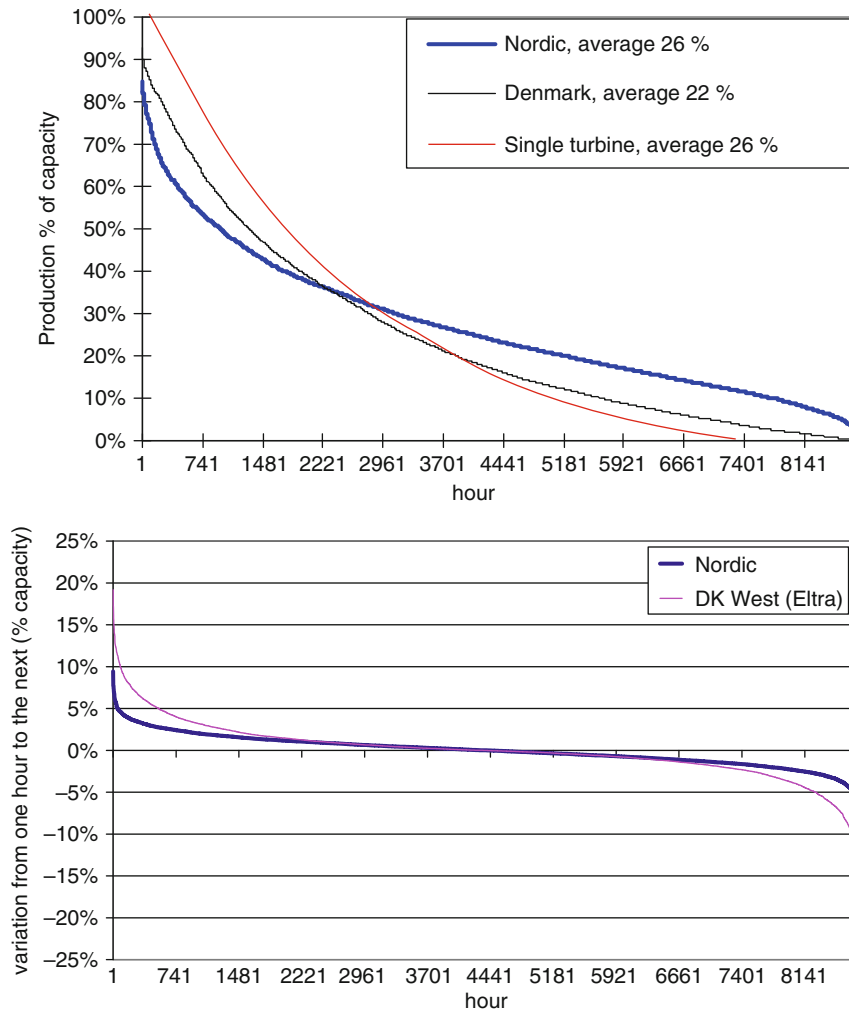
Generally, the variability of wind decreases as there are more turbines and wind power plants distributed over the area. Larger areas also decrease the number of hours of zero output – one wind power plant can have zero output for more than 1,000 h during a year, whereas the output of aggregated wind power in a very large area is always above 0 (Fig. 10). The variability also decreases as the time scale decreases – the second and minute variability of large-scale wind power is generally small, whereas the variability over several hours can be large even for distributed wind power. For time scales from several hours to day ahead, forecasting of wind power production is crucial.

Even if some general conclusions can be drawn from the variability of large-scale wind power, however, it should be noted that the size of the area and the way wind power plants are distributed is crucial. Also, the landscape can have influence. Offshore, the wind resource has been found to be more coherent, thus increasing the variability compared to similarly distributed wind power onshore. For power system operation, it is important to know how often there are higher



Wind Power Balancing. Figure 9

Wind power production (2,400 MW wind power) and load in Western Denmark. The worst storm event so far, of 8th January 2005, can be seen in hours 128–139 (Data source <http://www.energinet.dk>)



Wind Power Balancing. Figure 10

Wind power production and hourly variability as duration curves showing the smoothing effect from a larger area aggregated wind power [11]

variations in output, and also an estimation of the extreme event is important. For large-scale wind power where there are a high number of individual turbines and wind farms, the intermittency that may be seen in a single turbine output will become a more smooth variability visible in the hourly time scale.

General findings on large-scale variability can be summarized as [16]:

Very fast variations of distributed wind power are low (second-minute level). Even for a single wind power plant, the standard deviation of 1 s variations is only 0.1% of installed capacity. The 1 min variations (step changes) has the standard deviation from 2.1% to

0.6% of nominal capacity for a smaller and larger (250 turbines) wind farm [12]. There is increase in variability from the 10 min to the hourly time scale. The hourly variations do not smooth out very much inside one wind power plant.

The largest hourly step changes recorded from regional distributed wind power range from $\pm 10\%$ to $\pm 35\%$ depending on region size and how dispersed the wind power plants are. These are extreme values. Most of the time the hourly variations will be within $\pm 5\%$ of installed capacity [8]. The German example illustrates this: wind power changes are inside $\pm 1\%$ of the installed power 84% of time for 15 min

intervals and 70% of the time for 1 h time intervals (Fig. 11) [17].

Storm events can cause the most dramatic ramps – in West Denmark, for example, 80% of wind power production was lost during 6 h when wind speeds exceeded the cut-out limit of the turbines. Developing the forecasts for storm events is important.

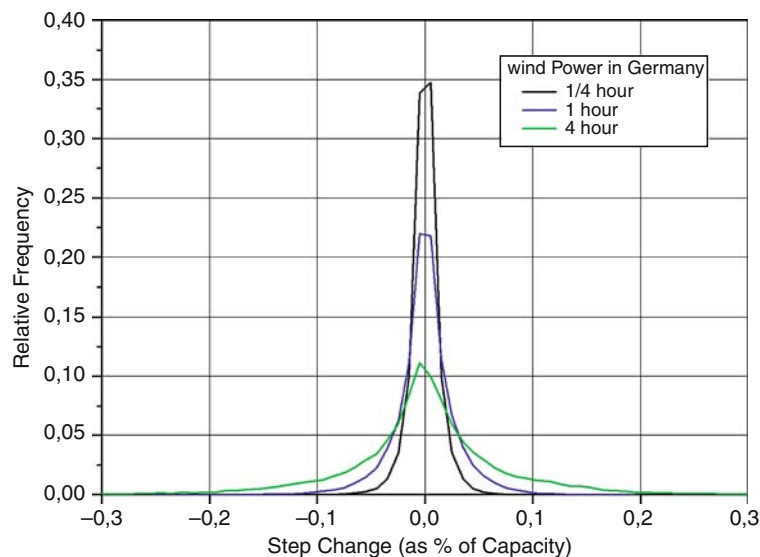
For large offshore wind power plants ramp rates can be more dramatic and this should be taken into account if most of the wind power capacity in the region is concentrated on one offshore site.

Forecast Errors of Distributed Wind Power Production The short-term forecasting of wind power production is still a recent power system tool when compared to load forecasting. For wind power, the level of accuracy will not be as high as for load. The experience so far shows that the overall shape of the production can be predicted most of the time. However, large deviations can occur both in the level and in the timing of the winds [18, 19]. For power system operation, it is also relevant to know the uncertainty of the forecast. At times of low uncertainty, there is not so much need for short-term reserve allocation and at times of high uncertainty reserve allocation and

potentially also measures of controlling wind power output can be used in system operation.

Level of accuracy improves when combining predictions for larger areas. For a single wind power plant the mean error for day-ahead forecasts is between 10% and 20% (as root-mean-square-error RMSE, % of nominal capacity). For a single balancing area this error will be below 10%. The latest results from West Denmark day-ahead forecasts show an average prediction error MAE (mean absolute error) of 6.0% of installed capacity, as an RMSE (root-mean-square-error) the error is 8.9%. In these numbers the relative forecast errors are to nominal capacity of wind power. When looking at the relative errors to average power (which give errors in terms of energy) the 6.0% for West Denmark corresponds to an error of 24% of yearly energy. The forecasts refer to day ahead, that is, $t + 13 \dots t + 37$ h and are based on a wind power capacity of 1,512 MW [8].

Further reductions can be expected from combining different forecasting models: The first results from Germany show the best model performing at 5.1% RMSE, a “simple” combination 4.2% and “intelligent” combination 3.9% [20]. An important issue is that when wind power is spread over large areas, the actual



Wind Power Balancing. Figure 11

Short-term changes within an hour of total wind power production [17]. A positive value reflects an increase in power and a negative value a decrease

production of other wind power plants in the up-wind direction can be used as one input for the forecast of the production in other wind power plants and thus for example a storm front crossing an area can be better forecasted some hours in advance.

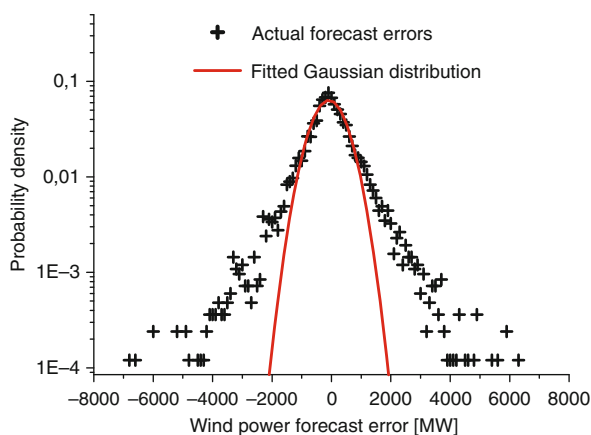
While the total balancing energy needed for the integration of wind power stems from the mean forecast error, the need for reserve power is closely connected to the largest forecast errors, that is, the tail in the probability density function (pdf) of forecast errors. Large wind power forecast errors are mainly caused by errors in the underlying weather prediction. Due to the chaotic nature of the weather, the pdf of the forecast error is not Gaussian [20]. Figure 12 shows as an example the probability density distribution of errors for a day-ahead wind power forecast for Germany; also shown is a fitted Gaussian distribution. It can be seen that large errors occur much more frequently than expected by a Gaussian distribution, causing a need for large reserve power in comparison to the balancing energy.

Main Challenges for Balancing for Power System Integrating Wind Power

There are several challenges regarding the handling of large amounts of wind power in power systems. The overall general challenge is to maintain a continuous balance, in an economical and reliable way

with minimum environmental effects. Within this context, there are two important special cases that are often discussed: low wind and high load situation and high wind in low load situation. The main challenge C1 and the two important special cases C2–C3 are described as following:

- C1. *Handling the continuous balancing.* The challenge for the continuous balancing process is that controllable production and consumption have to maintain the balance on all time scales. With larger amounts of wind power, there is a larger uncertainty because wind power forecasts are not as exact as the corresponding power consumption forecasts. For economic calculations the efficiency of power stations operating on part-load in order to compensate changes in production levels has to be taken into account. The continuous balancing process applies to both the technical capacity available for increasing other generation when the net load increases (wind power will make this harder if it decreases while the consumption is increasing) and for decreasing other generation in an economically feasible way when the net load decreases (wind power will make this harder during the instants when wind power increases while consumption is decreasing). This issue is further explained in the subsequent part of this section.
- C2. *Low wind and high load.* Situations with high power consumption are the basis for dimensioning the power generation capacity that has to be installed in a power system, often called as “capacity adequacy issue.” This challenging situation occurs comparatively seldom; every 5–10 years electricity consumption is very high. And power consumption may be high in the neighboring countries/regions, too, which reduces the possibilities of importing power. With large-scale wind power, this situation will persist, but not only on a consumption level, but also regarding variable wind power production. Looking at past situations of high power consumption in Sweden and Finland [22], there are situations both with high wind power production and with small amounts of wind power. When dimensioning the power system, wind power production both in the balancing area and in its neighboring areas have



Wind Power Balancing. Figure 12

Probability density distribution of errors for the day-ahead wind power forecast for Germany; also shown is a fitted Gaussian distribution [21]

to be taken into account regarding high power consumption levels. There is no power station with a 100% certain availability during times of high power consumption. Similarly to the fact that there is no guarantee that a thermal power station will work exactly at the time when power consumption peaks, you cannot assume that there are strong winds when power consumption is high. In order to arrive at an assessment whether the number of other power stations can be reduced due to wind power plants, the capacity value of wind power has to be estimated. This is done exactly the same way it is done for all other types of power stations, independent of whether they are wind, coal-fired, nuclear or hydro-power stations. The capacity value of wind power is often in the order of 10–20% of the installed wind power capacity. This issue is further discussed in chapter [Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit](#).

- C3. *High wind power and low load.* If you add large amounts of wind power to a certain area without changing anything else, this may result in very low prices and/or that not all wind power can be used and/or that thermal power stations often are used at low efficiency and/or that water has to be spilled past the turbines. If such situations occur often, it becomes economically beneficial to (a) increase transmission capacity from the area, (b) invest in pump storage in rivers (low price for pumping up to higher reservoirs and the water can be used when prices are higher), (c) have a flexible power consumption (e.g., power storage in district heating or charge electric vehicles during such situations). (d) further develop the low load production possibilities of other power plants. Some power plants cannot operate at lower than, for example, 60% of rated capacity but have to shut down. This is one example that Denmark has done to ease the wind power integration – decreasing the lowest possible generation level of thermal plants to be able to stay on line (and produce ancillary services). The key point here is that an appropriate dimensioning is an economic issue as low prices promote investment in suggestions (a)–(d); at the same time, large investments will increase electricity prices. It is difficult to actually dimension the

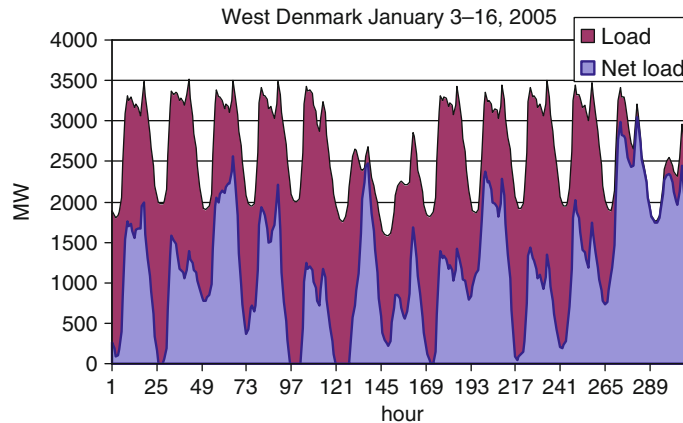
system “correctly” as lead times for investments cause an uncertainty regarding the investment planning. The “correct” dimensioning means in the practice that further investment (pump storage, transmission system, flexible power consumption) has a higher cost than the value it adds. It is important to remember that even a “correctly” dimensioned power system may include spill as it is not rational to dimension a power system without any spill. The conclusion is that – independent of whether the dimensioning is “correct” or not – it is very valuable that there is flexibility in both production and consumption, for example, that other production than wind power can produce at the lowest level possible.

How Wind Changes the Load Following on a Day-Ahead Time Scale

The scheduling process in power systems are made usually day ahead, as described in the previous section. Wind power forecasts are used in this process, in a similar way as load forecasts. The varying production patterns of wind generation changes the scheduling and unit commitment of the other production plants and use of transmission between regions. Either losses or benefits are introduced to the system, compared with the situation without wind. Most of the energy produced by wind power will impact the scheduling process, but there is a significant impact on the operating reserves as well, as the forecast errors of wind power can be large.

As there can be a large difference on the wind power production levels in daily time scale, there will be a large impact on how the conventional power plants are run in a power system with more than 10% wind power penetration (of yearly energy). Low wind and high wind days can mean a difference on conventional power production almost as much as the installed capacity of wind power. This is illustrated in [Fig. 13](#) by the West Denmark data from January, 2005 (wind penetration is about 24% of total gross demand and installed wind power capacity about 2,400 MW).

Large variations in wind power output can result in operating conventional power plants less efficiently. Optimized unit commitment, that is, planning the starts and shutdowns of slow-start units, is more



Wind Power Balancing. Figure 13

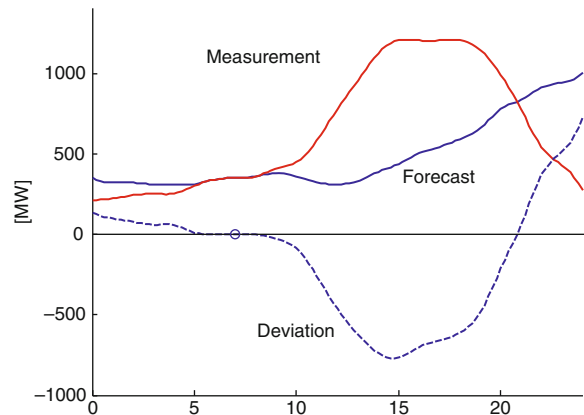
There will be a large impact of wind power on how the conventional power plants are run with a high wind penetration level (2 weeks from West Denmark, where wind power produces 24% of yearly electrical energy). Net load is electricity consumption (load) minus wind power generation

complicated when the variable and uncertain output from wind power is included. There will be more part load operation of conventional power plants. Start-ups and shut-downs of conventional power plants may also increase, especially to make room for the difference of low wind and high wind days. At higher penetration levels it has been noted that start-ups will decrease and more and more part-load operation at low levels is seen instead of shutting them down [11].

An important issue is also that the wind power forecasts are not perfect. This is illustrated in Fig. 14, where a real forecast is shown together with the real outcome.

It must though be noted that the large deviation considers a 24–32 h forecast since the drawn forecast is performed the day before at 11.00. For the actual scheduling of all hours, the flexibility of the existing system is essential. If, for example, all units can start up in 6 h, then the actual operation of the power plants in hour 13 can be planned at hour 7 which is illustrated with a circle in Fig. 11. This means that the 26 h forecast can be replaced with a 6 h forecast which is normally significantly better.

When wind power is added to the system, increase in costs due to increase in losses (increased operation in reduced efficiency) and wear and tear can be seen. This is however a moderate extra cost at least with penetration levels of 10–20% of wind power in the system.



Wind Power Balancing. Figure 14

Forecast, real production, and deviation for a day with large forecast error in the West Danish system

If wind power production exceeds the amount that can be absorbed while maintaining adequate reserve and dynamic control of the system, a part of the wind energy produced may need to be cut off (curtailed/spilled). It can though be noted that if 1% of wind power is spilled, then this increases the cost of the wind power with 1%. For a significant impact on the cost of wind power, the spillage has to be high. It can be noted that for, for example, hydro power investments, it is common to accept spillage in different situations, since this is found to be an economic solution.

For efficiency of production, the interest is on how the conventional capacity is run and how the variations and prediction errors of wind power change the unit commitment: both the time of operation and the way the units are operated (ramp rates, partial operation, starts/stops). These effects are in most cases analyzed by making simulations of system operation. In simulation models the reserve requirement is often taken into account, too, based on a simple rule how much reserves of moving up and down needs to be allocated to generation units in the simulation.

Wind power may influence the hydro power reservoir contents and reservoir management. When the wind production is large compared to the reservoir size in the area, there is a clear effect on reservoir management [11]. Wind power may also influence the losses of hydro power production. For example, a large wind power production in the spring flood time can result in the loss of some hydro production.

The impact of wind power on the power transmission depends on the location of wind power plants relative to the load, and the correlation between wind power production and load consumption. Wind power, like any load or generation, affects the power flow in the network and may even change the power flow direction in parts of the network. The changes in use of the power lines can bring about power losses or benefits. Increasing wind power production can affect bottleneck situations.

Modeling Wind Power in Dispatch The simulation results give insight into the technical impacts that wind power has, and also the (technical) costs involved. Analyses on how current market mechanisms affect wind power producers is also important. For wind power impact on the operation of power systems, modeling on an hourly level would be needed to catch the variability of wind. The models used in integration studies are often not specifically designed for dealing with high shares of wind power.

There are two critical challenges in energy system simulations: how the export possibilities to areas outside the simulated region are taken into account and how the wind power uncertainty at different time scales is taken into account in the model.

The effects of wind power on a power system are spread over the total control area (synchronously

operated area), balancing area or electricity market, with constraints on transmission capacities between the areas. Effects of wind power are first seen as an increase in exchange of electricity between the areas. The results will, however, show only increase in transmission unless boundary conditions are set, to be able to see the limits to the transfer of wind power variability to the neighboring areas. Contingencies, which have an impact on dynamic phenomena in the time scale of seconds, cannot be modeled with an hourly time scale model. Due to this, the low transmission possibility scenario is often used, and this can underestimate the exchange possibilities during most of the time.

In real life some corrections to schedules after day ahead are possible. Developing methods of incorporating wind power uncertainties into existing planning tools and models is of crucial importance. Modeling the effect of prediction errors is complicated by the uncertainties to different time scales of unit commitment (starting and shutting down slow thermal units) and dispatch (production levels of thermal units). Problems can also emerge from the simulation logic in itself: optimization of the system can be fundamentally different if taking into account the different nature of wind power production. Also, regulation requirements are often not modeled directly, but based on years of operating experience, so the effect of wind power cannot be modeled either [23]. Electricity market interactions and price levels should also be looked at when simulating wind power in the system.

Also the simulation case set up and comparison, to catch the effects of wind power, is of importance. Usually wind is added to the system and any effects are analyzed comparing the production and costs of the system with and without wind. For assessing the cost of variability of wind, the comparison can be made by adding wind as a flat production block over 24 h, or with a foreseeable diurnal pattern.

Which Power Production Will Wind Power Replace?

As has been explained above, production and consumption always balance. This means that all power that is fed into the power system at some point will always be used. If a wind power plant produces power, the need for other power production decreases accordingly. For each kilowatt hour of wind power, 1 kWh less is needed from another power plant (if the losses remain the same).

The question is which power stations will be replaced. Both in regulated and deregulated electricity systems, the use of the production form with highest marginal cost will be lowered by wind energy. But this may be different during different hours of the year. For most systems with some coal fired production plants, it is mostly coal power production that will be replaced by wind power. This is true when wind energy provides a minor amount of the total electricity consumption. This is also valid for large amounts of wind, for the countries where electricity production is based on coal. For other conditions, wind energy may replace gas fired production (with lower CO₂ emissions), or even CO₂ free production forms such as hydro, biomass or nuclear power. Even if the hydro production is reduced by wind energy, the hydro power stored in the reservoirs may be used later, possibly reducing fossil fuel fired production. Interconnected systems can also respond in such a way that wind power is partly replacing coal fired production in a neighboring country [11].

The exact result depends on the amount of wind power in the power system, and on the amount and costs of coal and gas fired production in the system.

In the following, some reasoning regarding the different power generation types and their role in power systems with large-scale wind power is provided:

Hydro-power: When wind power increases, it is often reasonable to decrease hydro-power production as it is comparatively easy to control. In practice, the amount of water passing through the power station is reduced and the water is stored in the dam instead. You have to keep in mind, though, that there will be more water in the reservoir that can be used later. In principle, you can say that for each kilowatt hour of wind power, it is possible to store 1 kWh of hydro-power in the dam. That means that wind power does *not* replace hydro-power, but shifts hydro-power production between hours. The amount of energy that hydro-power produces is mainly dependent on how much it rains. It is important, though, to remember that this only describes the general issue. There are many important details that have to be taken into account for an exact assessment of the opportunities. These include:

- During spring flood, the storage capacity of the dams may be limited.

- Different dropping heights and storage levels result in different efficiencies of hydro-power stations, that is, different amounts of energy per cubic meter of water.
- Forecast security requires certain margins as water that is dropped through the turbine of a power station will take, for instance, an hour to arrive at the next reservoir, and the conditions may have already changed by then.
- Water legislation sets clear limits to how reservoirs may be filled/emptied and how the plants may be operated. Similarly, power companies may even have self-imposed restrictions, that is, they do not make full use of the water legislation as this would lead, for instance, to inconveniences for local residents or tourists.
- It is a combination of water inflow, physical control options and water legislation that limits the control options in hydro-power stations.

Nuclear power: Nuclear power has an operating cost (cent per kilowatt hour) that is almost always lower than the electricity price. Therefore, they always try to produce as much as possible in these power stations. If wind power production increases, there are other power stations than nuclear that will decrease production instead. Nuclear generation can be reduced at times when there is a limited export options to countries with more expensive power production and in a situation where there is high wind power production and low load and all other generation is already cut-off. Nuclear power plants often have their yearly maintenance scheduled at times of spring floods when reduction of hydro power generation would result in spillage.

Coal/gas/oil power: There is coal power in the Scandinavian and Northern European power system that can be replaced by other power production. If it rains heavily in Sweden and/or there is a lot of wind power, it is mainly coal power that will reduce production in these countries because Sweden either exports power to these countries or reduces imports. The reason for reducing just coal power is that it has most often the highest operating costs. Sometimes, there are also other power stations, such as gas or oil-fired power plants, with operating costs that are even higher than that of coal power; and consequently, their production will be

decreased instead. It should be added that the green certificate trading leads to a situation where coal power prices increase the most (most carbon dioxide per kWh of produced power) which results in an increased interest in reducing just that production. One special case is combined heat and power production (CHP) where heat demand is setting the level of electricity production from the plants. However, in many cases there is a heat storage and thus also the electricity generation from these plants can technically be operated in a flexible way.

Conclusions: Wind power production results in that other power stations are operated at a lower capacity in comparison to a situation without wind power. If there are large amounts of hydro-power, then wind power will lead to the rescheduling of hydro-power to later points in time. Therefore wind power production at one point in time can indirectly lead to a reduced thermal power production at a later point in time. Wind power production mainly results in a reduced production in coal-fired power plants either in the area or in the neighboring areas. That requires a sufficient transmission capacity to countries with coal power, though, and that the actual pricing makes coal power the power with the highest operating costs.

How Wind Impacts the Balancing During the Operating Hour, Operating Reserves

Depending on the prediction accuracy, parts of the wind power production variations remain unpredicted or are predicted incorrectly, and there will also remain subhourly changes in output. This is what has to be handled by load following reserve or the balancing market.

During the hour of operation, the balance is first secured by means of primary reserve (automatic frequency control). In the event of major frequency deviation, the TSOs adjust the production or the consumption manually, using secondary/tertiary reserve. The reserve can be allocated by contracts between some producers (and consumers) and system operators or through a common regulating power market, where the players submit their bids for upward and downward regulation of production or consumption.

To estimate the impact of wind power on power system operating reserves, it has to be studied on

a control area basis. Every change in wind output does not need to be matched one-for-one by a change in another generating unit moving in the opposite direction. It is the total system aggregation, from all production units and consumption, which has to be balanced. A producer with flexible production could in principle counteract changes in its own wind power production also during the operational hour. Self regulation is discouraged, however, as it is more cost effective for both the system and the individual players to bid all regulating power to a joint pool for the system operator to use the cheapest options first. Otherwise there is a risk that, for example, wind power changes for one actor that physically are balanced out by changes in other wind power stations owned by another actor instead are balanced by each actor, which is not rational.

The increase in short-term reserve requirement is mostly estimated by statistical methods combining the variability or forecast errors of wind power to that of load and investigating the increase in the largest variations seen by the system [6]. Usually margins for “possible” extreme situations are kept. The term “possible” then normally includes a certain percentage of what in reality could happen. A straightforward method to define “possible” is the commonly used “ $N - 1$ criterion,” that is, it is necessary to keep reserves for an outage of the largest production unit or interconnector, which could be a challenge during both, import and export situations. This is a common dimensioning criterion for disturbance (contingency) reserve. In addition to that, some operational reserve is carried on top of that to cover variability and forecast errors, but there are no commonly used criteria to dimension this part of reserve. Some TSOs begin to use probabilistic approaches to define a suitable level of reserves.

Although wind power plants can at times create large “ramps” (i.e., large swings in production level), at current wind penetrations in most parts of the world these are not big enough or fast enough to be considered contingency events [5, 24].

Primary control is performed on a time scale of seconds/min. On this time scale, variations are already smoothed by different gusts for the individual turbines, inertia of the large rotors as well as variable speed turbines absorbing the variations, and there is no correlation between the variations of geographically dispersed wind farms. This is why the effect of wind power

on the system operation in the primary control time scale is very small even at considerable penetration [9, 25, 26].

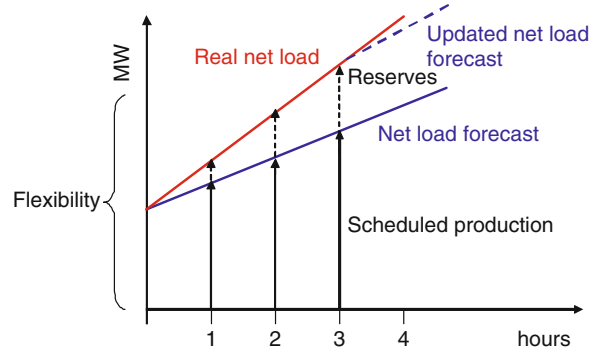
For operational reserves, the unforeseen variations induced from wind power are relevant on the time scale of $10 \text{ min}^{-1} \text{ h}$, where impacts of wind power can be clearly seen.

Methods to Estimate Allocation of Reserves Effects of wind power on reserve requirements, on a time scale of minutes, are often estimated based on statistical methods. The statistical approach for estimating the increase in reserve requirements is based on looking at the variability as a probability density. Combining the variability of wind with load variations, and looking at the increase in the net load variations is often referred to as the “ 3σ method” [27, 28]. This means that three times the standard deviation can be taken as a confidence level for how much of the variations should be covered by reserves (values of 2–7 have been used instead of 3; [6]).

It is of central importance to separate need of flexibility in longer time scales of several hours to a day (power plants that can follow net load variation) and need of reserves that can be activated in seconds or minutes time scale (power plants that can follow unforecasted net load variations). To illustrate the need of flexibility, Fig. 15 shows the needed power increase divided in scheduled production and reserves.

3 Sigma Methodology A widely used method for estimating the incremental increase of variability that the power system must balance when adding wind power is the consideration of the difference between the distribution of variations before and after wind power. As an estimate of the increase in variability, the standard deviation (σ) of the distribution can be used as a confidence level. Even if this is not the method by which the operating reserves are allocated in real power systems, this method is fairly easy to use and produces a value that is related to the degree by which wind power increases the variability in the power system [29].

Also the risk of forced outages can be included when estimating the increase in reserve requirements, which means combining the uncertainty of load, wind and other production. Because of spatial variations of



Wind Power Balancing. Figure 15

Flexibility in the power system to follow varying load and wind power production (the net load) consists of scheduled production and reserves. The need of flexibility is not the same as need of reserves, since a part of the net load variation can be forecasted

wind from turbine to turbine in a wind plant – and to a greater degree from plant to plant – a sudden loss of all wind power on a system simultaneously due to a loss of wind is not a dimensioning event. This is an important consideration for first contingency evaluation (disturbance/contingency reserves).

The reserve issue can be approached by noting that the system operator must balance the aggregate loads and generation. It is then possible to study the variability of load and wind power with time series of variations of wind power (P) and load (L):

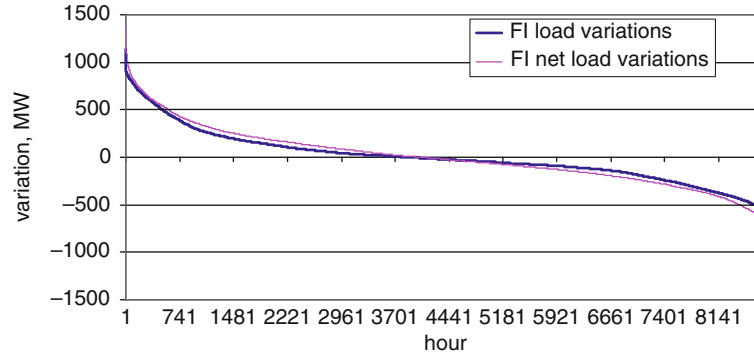
$$\Delta P_i = P_i - P_{i-1} \quad (1)$$

$$\Delta L_i = L_i - L_{i-1} \quad (2)$$

where i denotes the hour (from 2 to 8,760 in 1 non-leap year). The increase of variability that wind power brings to the power system can be seen when comparing their combination (net load) with the original load time series. Net load is the load minus the wind power production for each hour. The net load hourly variations are calculated like the hourly variations (1), but now for the net load time series (NL), where wind power production is subtracted from load:

$$\begin{aligned} \Delta NL_i &= NL_i - NL_{i-1} = (L_i - P_i) \\ &\quad - (L_{i-1} - P_{i-1}) = \Delta L_i - \Delta P_i \end{aligned} \quad (3)$$

In Fig. 16, the amount of hourly variations that the system sees is depicted showing 1 year of data in



Wind Power Balancing. Figure 16

Duration curves of annual load variations without wind power and net load variations with wind power. The case is for Finland in the year 2000 with hypothetical 6,000 MW wind (17% of gross demand) [11]

decreasing order (duration curve). There are two curves, the load variations without wind and load variations together with wind (the hourly variations of net load). The difference in the maximum values of MW of variations with and without wind indicates the amount that the operating reserve capacity must be increased. The difference in overall MW variations shown in the duration curves indicates the increased amount of existing reserve capacity operating, when wind power is added. The same capacity can in principle be used for both up and down regulation, and the variations as well as the increase should essentially be symmetrical. Either up or down (positive or negative) variations can determine a need for an increase in the reserves.

Planning and operating a power system is based on probabilities and risk. Reserves in the power system – with or without wind power – are generally determined so that variability within a certain probability are covered, for example 99.99% of the variability.

Standard deviation σ tells about the variability of the time series, it is the average deviation from mean value μ :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \quad (4)$$

For a normally distributed probability distribution, the standard deviation σ is a measure indicating that about 68% of the data is inside $\pm\sigma$ of the mean value. Taking a range of $\pm 3\sigma$ will cover 99%, and $\pm 4\sigma$ will

cover 99.99% of all variability. For hourly variations, the mean value is 0 by definition.

If the variability of load and wind power production can be assumed to be uncorrelated, the standard deviation of net load time series (σ_{NL}) can be determined by a simple square root sum of the standard deviations of load (σ_L) and wind power (σ_W) time series:

$$\sigma_{NL} = \sqrt{\sigma_L^2 + \sigma_W^2} \quad (5)$$

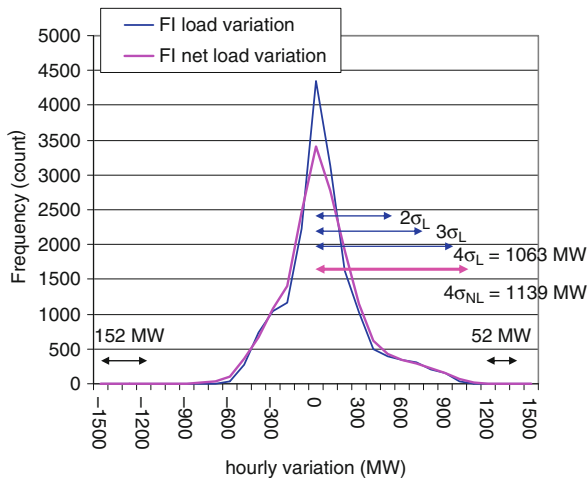
The hourly variability of wind power production and load do experience some correlation in the data of some regions. Also, the distribution of the variations is not usually strictly a normal distribution. However, formula (5) has been checked for several sets of data and it has produced accurate results for the standard deviation of the net load [30].

Finally, the increase, I , in the variability due to the wind can be formulated as the increase in some multiple of σ , calculating the increase from σ_L to σ_{NL} using for example 4σ as the confidence level covering most of the variability:

$$I = 4(\sigma_{NL} - \sigma_L) \quad (6)$$

In Fig. 17 the distribution curves for the variations in load and net load are depicted, showing also how far 2σ , 3σ , and 4σ reach to cover the variability of the dataset.

Exceedence Level A related approach to calculating load following reserve requirements for wind is to work the previous example backward. Instead of calculating some multiple of σ , which corresponds to a



Wind Power Balancing. Figure 17

An example of estimating the increase in hourly variability seen by the system ([31] Finland 2000–2001 data). If only maximum variation is looked at, the increase is determined at the tails of the distribution (52 MW increase in positive up-variation and 152 MW increase in negative down-variation). Looking at the standard deviation of the distributions, there is a difference of 76 MW in the 4σ coverage of the variability (1,139–1,063 MW)

given (low) probability that load will exceed generation, the probability level is chosen based on a priori risk tolerance. For example, if the chosen exceedence level is 99%, then the load (or load minus wind) can be sorted and the entry for the 99th percentile can be extracted. In a normal distribution, this would correspond to 3σ . However, if the underlying distribution is not normal, using an exceedence level can provide an alternative approach that provides the same level of risk.

To examine the reserve impact of wind, this approach can be easily done by (1) calculating a given exceedence level (percentile) for load alone, (2) calculate the same exceedence level (percentile) for the load net wind, and (3) comparing the incremental capacity of the two calculations.

Different Time Scale Reserves There are different requirements for different time scales in operational reserves, and this can be reflected in the amount of sigmas needed as a confidence level.

Regulating reserve must be available through the automatic generation control (AGC) scheme and

responds over time periods of up to several minutes. It is not uncommon for the system operator to carry regulating reserves equal to $\pm 5\sigma$ or $\pm 6\sigma$, where σ is the system variability measured as standard deviation of the load variability.

However, over longer time scales it may be possible to obtain generator responses from other sources, including non-AGC units that are spinning and synchronized, market participants, or quick-start generation that is not spinning. In systems with significant run-of-river hydro, the overall system variability will also be a function of the hydro variability over the time scale of 10–60 min, in addition to load and wind forecast errors and sub-hourly ramping. Some studies in the USA, for example, [31–33], have used 2σ as the preferred metric to calculate load following requirements for wind. This is based on the requirement (imposed by the North American Electricity Reliability Corporation, NERC) that the minimum required score for control performance standard 2 (CPS2) is 90%, which approximately corresponds to the normal probability value for 2σ . Other US studies have used $\pm 3\sigma$ as the appropriate confidence interval [24, 29, 34]. Another study undertaken by GE Energy of the Ontario system used $\pm 3\sigma$ and used a corresponding exceedence analysis as a supplement [35].

Combining the Impacts of Different Time Scales

When estimating the changes in balancing and any increase in reserve requirements this is usually done for different time scales. However, there are many time-based interrelations. Primary regulation and – to a certain degree – secondary regulation maintain the continuous balance. In order for that to work, it is necessary to continuously plan and re-plan so that there will be sufficient margins for maintaining the balance between production and consumption in a reliable and economic way. It is not possible to simply add up several types of reserves as they are interrelated. The events that occur during a certain hour, for instance, are a part of what can happen during the 24 h-period, cf. Fig. 18.

The difference between flexibility $P_{kn}(t)$ and reserves $P_{kr}(t)$ is that $P_{kn}(t)$ sets the total flexible capacity that has to be available, while $P_{kr}(t)$ only includes the part of the flexible capacity that must be available

within the time t without being prescheduled. Prescheduling means that power plants with longer start-up times are included in flexibility $P_{kn}(t)$, because it is assumed that their start-up process is initiated before the actual operation hour such that they are online in time period $k+t$. The capability of both $P_{kn}(t)$ and $P_{kr}(t)$ are strongly connected to the ramp rates of available power plants and in the case of $P_{kn}(t)$ start-up times of available power plants. The examples in Fig. 18 are mainly on hourly bases, but common reserve categories are, for example, “primary reserves” $=P_{kr}(\text{around } 1 \text{ min})$ and “secondary reserves” $=P_{kr}(\text{around } 1\text{--}5 \text{ min})$.

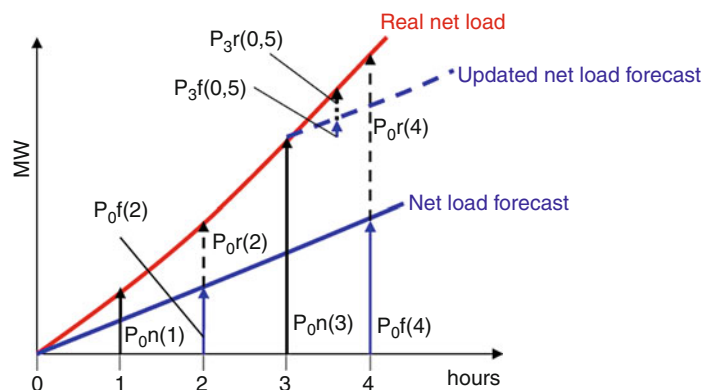
The examples in Fig. 18 show a case where production needs to be increased to follow the net load. There is also a need to decrease production when net load decreases. It is generally easier to decrease generation than to increase it, for example, by disconnection of generating plants (including wind power).

It can be noted that:

- Wind power can only decrease when there is a production level and only decrease largely when there is a large production. In a situation with no wind power, there is no need to include wind power when determining downward reserves or flexibility in the power system.
- In a situation with high wind power production, that is, a situation with possible decrease of wind

power production, other units are not in operation (since some of the load is met with wind power). This means that these units can increase their production if wind power decreases and/or load increases. This is based on the assumption that $P_{kn}(t)$ and $P_{kr}(t)$ can increase their production fast enough. This means that *flexibility* and *reserve keeping* in a system with wind power is often more an issue of ramp rates and start-up times, than a need of more capacity – more fast ramping and starting capacity can be needed, if the forecast errors are large enough that the slow units cannot follow.

- In many real systems (here called system A), there are connections to neighboring systems (system B). In system A a common approach is then to use a probabilistic (= statistical) approach concerning the need of reserves and flexibility, assuming that during a certain percent of the time the needed flexibility/reserves are imported from system B. But then the needed reserves in system A also have to consider possible outages in the transmission to system B. This means that one can estimate $P_{kn}(t)$ and $P_{kr}(t)$ from a probabilistic density function of possible changes (i.e., *variability* for P_{kn} or *forecast error* (for P_{kr}) and a certain percentage of accepted need for import of power. This method can also be used when one wants to make a rough estimate of needed $P_{kn}(t)$ or $P_{kr}(t)$ for a future system based on knowledge of $P_{kn}(t)$ and $P_{kr}(t)$ in the current



Wind Power Balancing. Figure 18

Need of “Flexibility” is the need for some units to follow the “net load variability,” marked $P_{kn}(t)$. This can be partly done by units scheduled beforehand, to follow the “net load forecast,” marked $P_{kf}(t)$, while “reserves” is the need for some units to follow the “net load forecast errors,” marked $P_{kr}(t)$

system. Also in systems without interconnections there are some dimensioning criteria which means that one only keep margins for “possible” situations. The term “possible” then normally includes a certain percentage of what in reality could happen. Also in these systems it is possible to use a probabilistic approach. A straightforward method to define “possible” is the commonly used “ $N-1$ criterion,” that is, it is necessary to keep reserves for an outage in the largest unit.

- The different time frames for both $P_{kn}(t)$ and $P_{kr}(t)$ are strongly interconnected. If one, for example, have enough $P_{kn}(4\text{ h})$, then this in general *includes* that during this period there must be enough $P_{kn}(1\text{ h})$. This means, in this case, that there is no meaning in calculating $P_{kn}(4\text{ h}) + P_{kn}(1\text{ h})$ since they overlap significantly. Such additions are only valid when they contain different units. In reality there is also an overlap between $P_{kn}(t)$ and $P_{kr}(t)$. Assume, for example, that $P_{kn}(t)$ is available, but the forecasted net load only increases $P_{kf}(t)$ (where $P_{kf}(t) < P_{kn}(t)$). This means that $P_{kn}(t) - P_{kf}(t)$ is available to be included in $P_{kr}(t)$. This is based on that this volume has a ramp rate that is high enough so it does not have to be prescheduled. It can also be noted that if the most extreme increase of net load that have ever happened is $P3(t)$, then this means that in this situation the highest requirement of *available flexibility* will be that $P_{kn}(t) = P3(t)$. It is then important to note that this includes $P_{kr}(t)$ so there is *not* a need for the system to cover both $P_{kn}(t) = P3(t)$ and a maximum $P_{kr}(t)$ to meet uncertainties *at the same time*.
- There is an interaction between “reserves” and market arrangements. If the net load forecasts are relatively accurate for a certain time frame (e.g., 24 h) then only a 24 h market is needed. If, on the other hand, the 24 h forecasts have low quality, then there has to be a market for changed production within that period. It is not the *variability* that requires updated markets but more the *net load forecast errors*. It must also be noted that, for example, “24 h reserves” is not a physical but mainly a market need. Most power plants need maximum 4–6 h to start up so they do not need this information 24 h in advance.

Wind Power in Electricity Markets

Wind power can take part in the electricity markets, where the energy can be traded according to wind power forecasts. As the accuracy of the forecasts is not as high as for other generation, or electricity consumption, there are almost always deviations from the traded amounts. If the imbalance prices are high, then intraday trading will considerably help the individual producers in the market to help lower the costs by better forecasts closer to delivery hour. Also aggregating wind power helps the actors in the market, as there is considerable decrease of (relative) forecast errors from dispersed production, just as is the case for load forecasts.

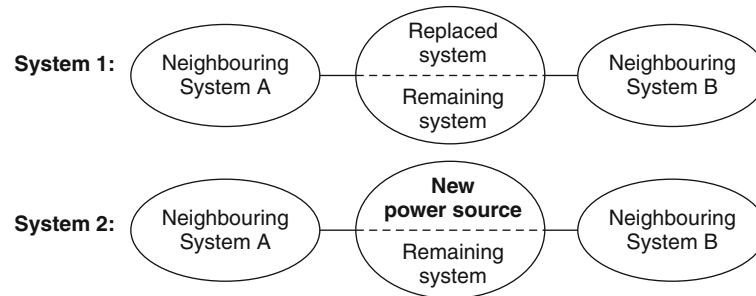
Balance settlement rules have a bit impact on the wind power producer. In systems where the imbalance prices are transparent and reflect the real system imbalance cost, wind power imbalance costs have been moderate [Holtinen et al., Neimane et al.]. In these systems wind power producers only pay extra costs during the hours when their imbalance has been to the same direction than system imbalance: for example need for upregulation when the whole system has used upregulation. When the imbalance costs are low, at low wind penetration levels, intraday trading is not always profitable for the actors. If the imbalance prices contain penalties for all deviations, no matter to which direction the system imbalance has been, then operating with shorter forecasts, intraday, is crucial.

Results from Integration Studies

An Integration study of wind power is often performed in order to obtain the possibilities to integrate a certain amount of wind power in a certain system and/or to estimate needed changes in the system set up in order to make it possible to integrate the wind power. “Integration cost” is then defined as the additional costs that are required in the system (except the costs for installing the new power plants) to keep customer requirement at an acceptable reliability, cf. Fig. 19. [36].

First two terms have to be defined:

- *The replaced system* = The power system consisting of other power sources in the system that will be replaced by the new power source



Wind Power Balancing. Figure 19

Integration cost = [costs in system 2] – [costs in system 1]

- *The remaining system* = The power system consisting of the other power sources, that are the same in both alternative systems

It is important to note that how the *replaced system* is selected is essential for the result. If the alternative to wind power (i.e., the *replaced system*) is nuclear power that is not taking part in regulation (as the amount of production changes should be kept small) the consequence of having wind power is also smaller compared to if wind power would replace easily regulated gas-fired generation.

Also the selection of the *remaining system* is essential. If wind power is to be integrated in a system with a large share of nuclear power (i.e., the *remaining system*) this will then cause a much larger “integration cost” compared to a study where the *remaining system* consists of easily regulated gas-fired generation.

Integration cost can be divided into different components arising from the increase in the operational balancing cost and grid expansion cost. The value of the capacity credit of wind power can also be stated. Here only the results on “balancing” are summarized.

Many wind integration studies give estimated impacts as increase in reserve requirements (MW) and integration costs (\$/MWh, €/MWh). Many studies give the results in less comparable ways, like impacts on the scheduling of other power plants and exports.

Different metrics for the results have been used in the studies: Results as monetary value per MWh of wind or per MWh of total consumption (reflecting the increase in consumer price). There are also results as % of more wind power production needed to cover extra losses.

For increased balancing it is important to note whether a market cost has been estimated or whether the results refer to technical cost for the power system. There is also benefit when adding wind power to power systems: it reduces the total operating costs and emissions as wind replaces fossil fuels. The trade-off between curtailing wind output in critical times and providing new transmission or production capacity would be needed in some cases 37.

In the following graphs only the cost component has been analyzed. The case studies summarized are not easy to compare due to different methodology and data used, as well as different assumptions on the interconnection capacity available.

Determining what is “high” penetration of wind power is not straightforward. Often either energy or capacity metrics are used: wind power production as % of gross demand (energy) and wind power as % of peak load (capacity). To determine high penetration for a power system also interconnecting capacity needs to be looked at. This is because critical moments of high wind and low load can be relieved by using interconnector capacity.

Results for Increase in Reserve Requirements

Wind generation may require system operators to carry additional operating reserves. From both the experience and results from studies performed, a significant challenge is the variability of wind power within 1–6 h. Frequency control (time scale of seconds) and inertial response are not crucial problems when integrating wind power into large systems at the present time, but may be a challenge for small systems (like Ireland)

and will become more of a challenge for systems with high penetration in the future.

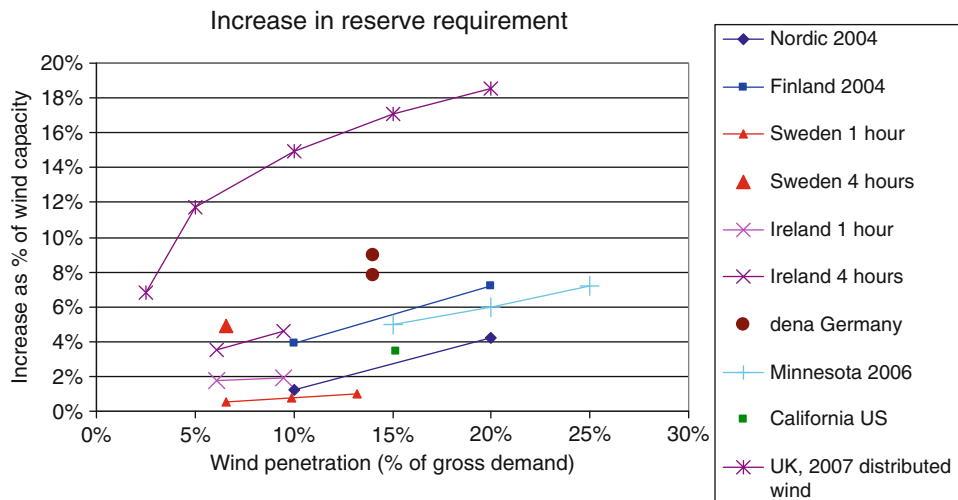
The results presented in Fig. 20 for increase in reserve requirements due to wind power are from following studies: Finland and Nordic [11], Sweden [37], Ireland [38], the UK [39], Germany [40]; Minnesota 2006 [32] and California [41].

The estimated increase in short-term reserve requirements in the studies has a large range: 1–15% of installed wind power capacity at 10% penetration (of gross demand) and 4–18% of installed wind power capacity at 20% penetration. Time scales used in the estimation explain much of the differences in results:

- If only hourly variability of wind is taken into account when estimating the increase in short-term reserve requirement, the results are 0.5–4% of installed wind capacity or less, with penetrations below 10% of gross demand.
- When 4 h forecast errors of wind power are taken into account, an increase in short-term reserve requirement of 4–5% of installed wind capacity has been reported, with penetration levels of 5–10% of gross demand.

The highest results in Fig. 20 are from a study where 4 h variability of wind (not forecast error), combined with load forecast error, results in 15% reserve requirement at 10% penetration and 18% reserve requirement at 20% penetration of gross demand [39]. The latest achievements in wind forecasting show a considerable improvement of predictions also in short time scales, so using updated forecasts would reduce this estimate. If day-ahead forecast errors are left to be balanced with the short-term reserves, the increase in short-term reserve requirement is nearly 10% (two points for Germany reflect difference in up- and down-regulation requirements) [40]. In this German study, the reserve requirement is taken as the average impact of day-ahead forecast errors of wind power. The maximum values would result in an increase that is 15–20% of installed wind capacity.

In most studies, reserve requirements for different time frames are summed up – for example, increase in regulating reserves + increase in load following reserves. There are some studies showing larger increase in reserve requirement than shown here. Swedish TSO published estimates that are 35–48% of installed wind capacity [42]. This is due to adding up several cases and time frames of reserves. However, the different time



Wind Power Balancing. Figure 20

Results for the increase in reserve requirement due to wind power. German Dena estimates are taking into account the day-ahead uncertainty (for up and down reserves separately) and UK the variability of wind 4 h ahead. In Minnesota and California, day-ahead uncertainty has been included in the estimate. For the others the effect of variations during the operating hour is considered. For Ireland and Sweden the 4 h-ahead uncertainty has been evaluated separately

frames, for example, 1 h ahead and 4 h ahead, are strongly overlapping. If one has enough flexibility for 4 h, then this in general implies that during this period there is enough flexibility for 1 h ahead. In this case there is no meaning in summing up these reserve requirements since they overlap significantly. Such additions are only valid when they contain different units.

An important issue is that “increase in reserve requirements” does not necessarily mean need of new investments. The amount of wind-caused reserves is at highest when wind power is on a high production level. In these situations the other power stations are operated on a low level, which means that they can act as reserves and increase the generation if wind power decreases. This means that flexibility and reserve keeping in a system with wind power is an issue of ramp rates and start-up times, together with a need of more capacity – more fast ramping and starting capacity can be needed, if the forecast errors are large enough that the slow units cannot follow. This must be considered when “increased reserve margins” are to be estimated.

Balancing Cost

Wind power impacts on power system balancing can be seen in several time scales, from minutes to hours, up to the day-ahead time scale. General conclusions on increase in balancing requirement will depend on region size relevant for balancing, initial load variations and how distributed wind power is sited. Here also, the operational routines of the power system are relevant – how often the forecasts of load and wind are updated, for example.

To arrive at estimates for balancing cost, the operating reserve impact is one issue (increase in reserve requirement from statistical methods) and impact on efficiency of conventional power plants for day-ahead operation is another issue (simulations). For the simulations most results are based on comparing costs of system operation without wind and adding different amounts of wind. The costs of variability are also addressed by comparing simulations with flat wind energy to varying wind energy, for example in Minnesota, USA [32] and Greennet Nordic+Germany [43].

It is important to pay attention to the representativeness of wind input data (how well does the wind

data represent the geospread of the power system, how is wind power simulated, what time scale effects on variability and predictability have been taken into account) and also how the main setup for the assessment or simulation is made (wind power replacing other production or capacity and to what extent is the power system operation optimized when wind power production is added). The level of detail of the simulation model (time resolution, level of detail in simulating conventional generation and transmission, pricing) and how the uncertainty in the wind plant output forecast is handled with respect to the load forecast uncertainty are also important.

The check-list developed in [35] can be used to find out whether the simulation approach has been conservative or whether some important aspects have been omitted, producing either high or low estimates for the impacts. The most general finding comparing the study setups is the use of interconnection capacity – this is crucial when estimating the impacts of wind power.

The results presented in Fig. 21 for increase in balancing costs due to wind power are from the following studies: Finland and Nordic countries [11], the UK [44, 39], Ireland [45], Colorado [45], Minnesota [46, 32], California [47, 48]; Nordic countries and Germany [43].

From the cost estimates presented in investigated studies it follows that at wind penetrations of up to 20% of gross demand (energy), system operating cost increases arising from wind variability and uncertainty amounted to about 1–4 €/MWh wind power produced (Fig. 21). This is 10% or less of the wholesale value of the wind energy. The actual impact of adding wind generation in different balancing areas can vary depending on local factors. Important factors identified to reduce integration costs are aggregating wind plant output over large geographical regions, larger balancing areas, and operating the power system closer to the delivery hour with accurate forecast systems.

The highest estimates of reserve requirements from Germany and the UK are not reflected in balancing costs, as from both studies it was concluded that this amount of reserve can be handled with the current conventional power plants. From the UK, only the increased cost of operating existing reserves has been estimated.

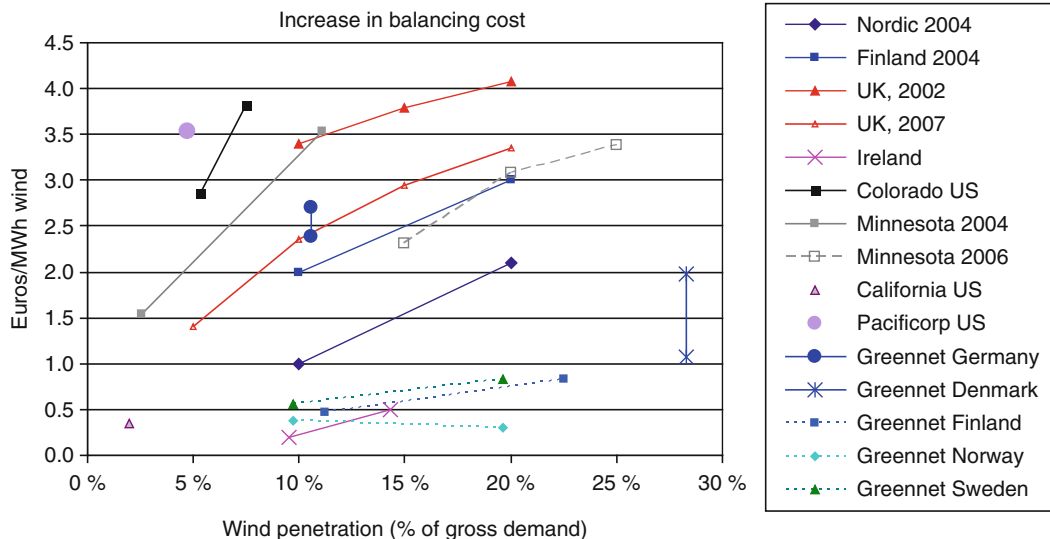
In addition to estimates, there is some experience from Denmark for the actual balancing costs for the existing wind power. For West Denmark, the balancing cost from the Nordic day-ahead market has been 1.4–2.6 €/MWh for a 24% wind penetration (of gross demand). These numbers are quite in the middle of theoretically estimated results from studies in Fig. 21. The balancing cost paid by the wind power producers is not the same as increase in technical cost of balancing for the amount of wind power. However the market costs from Denmark and Spain reflect the cost incurred to the system. In Denmark the cost is low compared to the penetration level, as the cost comes from Nordic market where wind power penetration is still small.

The interconnection capacity to neighboring systems is often significant. For the balancing costs, it is then essential to note in the study setup whether the interconnection capacity can be used for balancing purposes or not. A general conclusion is that if interconnection capacity is allowed to be used also for balancing purposes, then the balancing costs are lower compared to the case where they are not allowed to be used.

Balancing Cost from Electricity Markets In Finland and Sweden, the balancing costs as payments for wind

power producers have been estimated from the balancing market (Nordic Regulating market) prices to be 0.3–1.4 €/MWh depending on how distributed the wind power is and on the market price level for balancing [49, 50]. In Sweden, the use of 15 min operating reserves has been estimated to increase by 18–56% of current amounts due to wind power forecast errors 1 or 4 h ahead for 4,000 MW wind power (8% of gross demand) [51]. The increased cost of system imbalances of Finland due to future wind power prediction errors was estimated to be 0.2–1 €/MWh for penetration levels of 1–10% of gross demand, assuming the Nordic balancing market was available (no bottlenecks) [52].

The use of an intraday market to help reduce the imbalance costs of wind power has been examined in Germany [53] and for the Nordic market in Finland [54] and Sweden [51] have shown that for the current price assumptions there is not a straightforward benefit to use an intraday market. This is because trading at an intraday market would mean correcting all imbalances, whereas the imbalance payments only apply to the imbalances that affect the power system net imbalances, thus not 100% of time (at low wind penetrations only 50% of time).



Wind Power Balancing. Figure 21

Results from estimates for the increase in balancing and operating costs due to wind power. The currency conversion used here is 1 € = 0.7 £ and 1 € = 1.3 US\$. For UK, 2007 study the average cost is presented here, the range in the last point for 20% penetration level is from 2.6 to 4.7 €/MWh

Other Balancing Related Results

Not all case studies presented results quantified as MW of increase in reserve requirements or monetary values for increase in balancing costs.

In Denmark the TSO has estimated the impacts of increasing the wind penetration level from 20% to 50% (of gross demand) and concluded that further large-scale integration of wind power calls for exploiting both, domestic flexibility and international power markets with measures on the market side, production side, transmission side, and demand side [55, 56].

For the Netherlands, the simulations show the benefit of international trade of electricity and postponing market gate closure for wind integration. Wind power worsens the business case for thermal generation: CCGT during peak demand and base-load coal during low demand [57].

The Irish All Island Grid Study shows that going from 2 to 6 GW wind, the operational costs of the electricity system fall by €13/MWh when compared to the base case – due to cost benefit approach in the study, the cost component was not published as such [58].

For New York, 10% penetration of capacity, incremental regulation due to wind was found to be 36 MW. No additional spinning reserve was needed. Incremental intra-hour load following burden increased 1–2 MW/5 min. Hourly ramp increased from 858 MW to 910 MW. All increased needs can be met by existing NY resources and market processes. System cost savings of \$335–\$455 million for assumed 2008 natural gas prices of \$6.50–\$6.80/MMBTU were found. Day-ahead unit-commitment forecast error σ increased from 700–800 to 859–950 MW. Total system variable cost savings increases from \$335 million to \$430 million when state-of-the-art forecasting is considered in unit commitment (\$10.70/MWh of wind) [59].

Storage The value of storage in the power system operation in the UK was estimated to be 252–970 £/kW [37]. For Germany a 27 M€/year revenue could be foreseen for 400 MW CAES (250 M€ investment) [54]. In the NL international exchange was seen as a more promising alternative to storage in the system [58]. In Ireland adding storage did not bring additional value in the All Island Grid Study results [59].

For wind penetration levels of 10–20% of gross demand in power systems, the cost effectiveness of building new electricity storage is still low (excluding hydro power with large reservoirs or pumped hydro). With higher wind penetration levels the extra flexibility that also storages can provide will be beneficial for the power system operation, provided they are economically competitive with other forms of flexibility. It is important to notice, however, that any storage should be operated according to the needs of aggregated system balancing. It is not cost effective to provide dedicated back-up for wind power in large power systems where the variability of all loads and generators are effectively reduced by aggregating, in the same way as it is not effective to have dedicated storage for outages in a certain thermal power plant, or having specific plants following the variation of a certain load.

Conclusions and Future Directions

Wind power impacts on power system balancing can be seen in several time scales, from minutes to hours, up to the day-ahead time scale. General conclusions on increase in balancing requirement will depend on region size relevant for balancing, initial load variations and how concentrated/distributed wind power is sited.

Several issues that impact on the amount of wind power that can be integrated have been identified. Aggregation benefits of large areas help in reducing the variability and forecast errors of wind power as well as help in pooling more cost effective balancing resources. An alternative to large balancing areas is to allow and promote intraday and intra-hour trading between different balancing areas in order to obtain low-cost balancing services. System scheduling and operating electricity markets at less than day-ahead time scales help reduce the forecast errors of wind power that affect operating reserves. Transmission is the key to aggregation benefits, electricity markets and larger balancing areas.

Wind integration has mainly been studied to wind penetration levels of 10–20% of gross demand (up to 50% of peak load), with some first efforts to study higher penetration levels of 40–50% of gross demand [56, 57, 59]. What happens in larger penetration levels, where wind becomes a more dominating part of power system, is not completely clear. Studies will have to

cover larger areas to take the cross border transmission into account properly [60]. The future power systems may also provide different options for flexibility in demand side that do not exist today. Future integration studies should take into account the foreseen high penetration of PV or ocean power. This will help smoothing the variability of individual technologies.

Recommendations for integration studies:

It is very important to take the variability of wind into account in a right way in power system studies. The variability will smooth out to some extent if there is geospread wind power, and part of the variability can be forecast. Because of spatial variations of wind from turbine to turbine in a wind power plant – and to a greater degree from wind power plant to wind power plant – a sudden loss of all wind power on a system simultaneously due to a loss of wind is not a credible event. Sudden loss of large amounts of wind power due to voltage dips in the grid can be prevented by requiring fault-ride-through from the turbines. Recommendations for wind integration studies include:

- Capturing the smoothed out variability of wind power production time series for the geographic diversity assumed and utilizing wind forecasting best practice for the uncertainty of wind power production
- Examining wind variation in combination with load variations, coupled with actual historic utility load and load forecasts
- Capturing system characteristics and response through operational simulations and modeling
- Examining actual costs independent of tariff design structure
- Comparing the costs and benefits of wind power

In most cases the question is whether extra investments to power systems are economically profitable or not in the new system with larger amount of wind power – not only stating that a certain amount of extra reserve capacity and/or new transmission lines, etc., are a prerequisite in order to build any wind power.

For high penetration levels of wind power, the optimization of the integrated system should be explored. Modifications to system configuration and operation practices to accommodate high wind penetration may be required. Not all current system operation techniques are designed to correctly incorporate the characteristics of wind generation and surely were not

developed with that objective in mind. Increasing power system flexibility through such means as transmission to neighboring areas, generation flexibility, demand side management and optimal use of storage (e.g., pumping hydro or thermal) in combination with market aggregation and operation closer to real time will impact the amount of wind that can be integrated cost effectively.

Bibliography

Primary Literature

1. Söder L (2002) Explaining power system operation to nonengineers. *Power Eng Rev*, April 2002
2. Hirvonen R (2000) Material for course S-18.113 Sähköenergiajärjestelmät. Helsinki University of Technology Power systems laboratory, 2000 (in Finnish)
3. ETSO-2006 (2006) An overview of current cross-border congestion management methods in Europe, European Transmission System Operators, May 2006, available from <http://www.entsoe.eu/>
4. Kristoffersen JR (2005) The horns rev wind farm and the operational experience with the wind farm main controller. In: *Proceedings of Copenhagen Offshore Wind*, October 2005, Copenhagen, Denmark
5. Holttinen H, Milligan M, Kirby B, Acker T, Neimane V, Molinski T (2008) Using standard deviation as a measure of increased operational reserve requirement for wind power. *Wind Eng* 32(4):355–377
6. Milligan M, Porter K, DeMeo E, Denholm P, Holttinen H, Kirby B, Miller N, Mills A, O'Malley M, Schuerger M, Soder L (2009) Wind power myths debunked. *IEEE Power Energy Mag* 7(6):89–99
7. Soder L, Hofmann L, Orths A, Holttinen H, Wan Y, Tuohy A (2007) Experience from wind integration in some high penetration areas. *IEEE T Energy Convers* 22(1):4–12
8. Holttinen H, Meibom P, Orths A, van Hulle F, Lange B, O'Malley M, Pierik J, Ummels B, Tande JO, Estanqueiro A, Matos M, Gomez E, Söder L, Strbac G, Shakoor A, Ricardo J, Smith JC, Milligan M, Ela E (2009) Design and operation of power systems with large amounts of wind power. Final report, IEA WIND Task 25, Phase one 2006–2008. Espoo, VTT. 200 p. + app. 29 p. VTT Tiedotteita – Research Notes 2493. <http://www.vtt.fi/inf/pdf/tiedotteet/2009/T2493.pdf>
9. Ernst B (1999) Analysis of wind power ancillary services characteristics with German 250 MW wind data. NREL Report No. TP-500-26969. 38 p. <http://www.nrel.gov/publications/>
10. Focken U, Lange M, Waldl H-P (2001) Previento – a wind power prediction system with an innovative upscaling algorithm. In: *Proceedings of EWEC'01*, 2nd–6th July, 2001, Copenhagen, pp 826–829
11. Holttinen H (2004) The impact of large scale wind power production on the Nordic electricity system. PhD dissertation.

- VTT Publications 554. Espoo, VTT Processes. 82 p. + app. 111 p. <http://www.vtt.fi/inf/pdf/publications/2004/P554.pdf>
12. Wan Y (2005) Fluctuation and ramping characteristics of large wind power plants. Windpower 2005 (Windpower 05) Conference and Exhibition (CD-ROM), 15–18 May 2005, Denver, Colorado. Washington, DC: American Wind Energy Association; Content Management Corp. NREL Report No. CP-500-38057. 13 p
 13. EWEA (2005) Large scale integration of wind energy in the European power supply: analysis, issues and recommendations (December 2005). <http://www.ewea.org/>
 14. IEA (2005) Variability of wind power and other renewables. Management options and strategies. http://www.iea.org/Textbase/publications/free_new_Desc.asp?PUBS_ID=1572
 15. Giebel G (2007) A variance analysis of the capacity displaced by wind energy in Europe. Wind Energy 10:69–79
 16. Holttinen H, Meibom P, Orths A, Lange B, O'Malley M, Tande JO, Estanqueiro A, Gomez E, Söder L, Strbac G, Smith JC, van Hulle, F. (2010) Impacts of large amounts of wind power on design and operation of power systems, results of IEA collaboration. Submitted and accepted to Wind Energy journal
 17. ISET (2005) Wind energy report Germany 2005. ISET, Kassel
 18. Giebel G, Brownsword R, Kariniotakis G (2003) The state-of-the-art in short-term prediction of wind power. A literature overview. EU project ANEMOS (ENK5-CT-2002-00665). <http://anemos.cma.fr>
 19. Kariniotakis G et al (2006) Next generation forecasting tools for the optimal management of wind generation. In: Proceedings PMAPS Conference, Probabilistic Methods Applied to Power Systems, KTH, Stockholm, Sweden, June 2006
 20. Focken U (2007) Optimal combination of European weather models for improved wind power predictions. In: Proceedings of EWEC'07, 7th–10th May, 2007, Milan, Italy
 21. Lange B, Cali Ü, Jursa R, Mackensen R, Rohrig K, Schlögl F (2006) Strategies for balancing wind power in Germany. In: German Wind Energy Conference DEWEK 2006, Bremen, November 2006
 22. Holttinen H, Stenberg A (2008) Tuulivoiman tuotantotilastot. Vuosiraportti 2008. Espoo, VTT. 47 s. + liitti. 8 s. VTT Working Papers; 132. <http://www.vtt.fi/inf/pdf/workingpapers/2009/W132.pdf>
 23. Dragoon K, Milligan M (2003) Assessing wind integration costs with dispatch models: A case study. Windpower, Austin
 24. Milligan M, Kirby B (2007) The impact of balancing areas size, obligation sharing, and ramping capability. In: Proceedings of the WindPower 2007 Conference, June 3–5, 2007. Los Angeles, CA, USA. NREL-CP-500-41809. <http://www.nrel.gov/docs/fy07osti/41809.pdf>
 25. Kirby B, Milligan M, Makarov Y, Hawkins D, Jackson K, Shiu H (2003) California Renewables Portfolio Standard. Renewable generation integration cost analysis. Phase I: One year analysis of existing resources. Results and recommendations. California Energy Commission/California Public Utilities Commission. <http://cwec.ucdavis.edu/rpsintegration/>
 26. Dany G (2001) Power reserve in interconnected systems with high wind power production. In: IEEE Power Tech Conference, 10–15th September, 2001, Porto, Portugal
 27. Soder L (1993) Reserve margin planning in a wind-hydro-thermal power system. IEEE Trans Power Syst 8(2):564–571
 28. Milligan M (2003) Wind power plants and system operation in the hourly time domain. In: Proceedings of Windpower 2003 Conference. May 18–21, 2003 Austin, Texas, USA. NREL/CP-500-33955. <http://www.nrel.gov/publications/>
 29. Kirby B, Hirst E (2000) Customer-specific metrics for the regulation and load following ancillary services. ORNL/CON-474, Oak Ridge National Laboratory, Oak Ridge TN, January 2000
 30. Holttinen H (2005) Impact of hourly wind power variations on the system operation in the Nordic countries. Wind Energy 8(2):197–218
 31. EnerNex/WindLogics (2006) Minnesota Wind Integration Study Final Report. vol I, prepared for Minnesota Public Utilities Commission, November 2006. http://www.puc.state.mn.us/portal/groups/public/documents/pdf_files/000664.pdf
 32. EnerNex Corporation (2007) Final Report – Avista Corporation Wind Integration Study. Avista Utilities. <http://www.uwig.org/AvistaWindIntegrationStudy.pdf>
 33. Idaho Power Corporation (2007) Operational impacts of integrating wind generation into idaho power's existing resource portfolio. <http://www.idahopower.com/pdfs/energycenter/wind/windIntegrationstudy.pdf>
 34. GE Energy (2007) Californian intermittency analysis. <http://www.uwig.org/CEC-500-2007-081-APA.pdf>
 35. GE (2006) Ontario Wind Integration Study. Final Report to: Ontario Power Authority (OPA), Independent Electricity System Operator (IESO) and Canadian Wind Energy Association (CanWEA). <http://www.uwig.org/OPA-Report-200610-1.pdf>
 36. Soder L (2008) On methodology for modelling wind power impact on power systems (R. Inst. Technol., Stockholm, Sweden); Holttinen H (2008) International Journal of Global Energy Issues, 29(1–2):181–198
 37. Axelsson U, Murray R, Neimane V (2005) 4000 MW wind power in Sweden – Impact on regulation and reserve requirements. Elforsk Report 05:19, Stockholm, 2005. <http://www.elforsk.se>
 38. Ilex, UMIIST, UCD and QUB (2004) Operating reserve requirements as wind power penetration increases in the Irish electricity system. Sustainable Energy Ireland, 2004
 39. Strbac G, Shakoar A, Black M, Pudjianto D, Bopp T (2007) Impact of wind generation on the operation and development of the UK electricity systems. Electr Power Syst Res 77(9):1143–1238, Elsevier
 40. Planning of the grid integration of wind energy in Germany onshore and offshore up to the year 2020 (Dena Grid study). Deutsche Energie-Agentur Dena, March 2005. English summary and full German version <http://www.dena.de/themen/thema-reg/projektarchiv/>
 41. Porter et al (2007) Intermittency analysis Final report. July, 2007. <http://www.uwig.org/CEC-500-2007-081.pdf>
 42. Kraftnät S (2008) Large scale expansion of wind power – consequences for the transmission grid and need of

- regulation power. http://www.svk.se/Global/01_Om_oss/Pdf/Rapporter/080601_Bilaga_vindkraftsrapport_2008.pdf (In Swedish)
43. Meibom P, Weber C, Barth R, Brand H (2009) Operational costs induced by fluctuating wind power production in Germany and Scandinavia. *IET Renew Energy Gener* 3(1):75–83, March 2009
 44. Strbac G (2002) Ilex Energy. Quantifying the system costs of additional renewables in 2020. DTI, 2002. http://www.dti.gov.uk/energy/developpep/080scar_report_v2_0.pdf
 45. Zavadil R (2006) Wind Integration Study for Public Service Company of Colorado. May 22, 2006. http://www.xcelenergy.com/XLWEB/CDA/0,3080,1-1-1_1875_15056_15473-13518-2_171_258-0,00.html
 46. EnerNex/WindLogics (2004) Xcel North study (Minnesota Department of Commerce). <http://www.state.mn.us/cgi-bin/portal/mn/jsp/content.do?contentid=536904447&contenttype=EDITORIAL&hpage=true&agency=Commerce>
 47. Shiu H, Milligan M, Kirby B, Jackson K (2006) California Renewables Portfolio Standard Renewable Generation Integration Cost Analysis. California Energy Commission, PIER Public Interest Energy Research Programme. http://www.energy.ca.gov/pier/final_project_reports/CEC-500-2006-064.html
 48. PacifiCorp (2005) Integrated Resource Planning. <http://www.pacificcorp.com/Navigation/Navigation23807.html>
 49. Holttinen H, Saarikivi P, Repo S, Ikäheimo J, Koreneff G (2006) Prediction errors and balancing costs for wind power production in Finland. In: Proceedings of 6th workshop on Offshore and Large Scale Integration of Wind Power, 25–26th October, 2006, Delft, The Netherlands
 50. Neimane V, Carlsson F (2008) A massive introduction of wind power. Changed market conditions? Elforsk report 08:41. http://www.vindenergi.org/Vindforskrappporter/v_132.pdf
 51. Brandberg M, Broman N (2007) Future trading with regulating power, Magnus Brandberg and Niclas Broman, Master's Thesis, Uppsala Universitet. An updated version, together with Nilsson, in Minerals and Energy – Raw Materials Report, vol 23(1): March 2008, pp 1–11
 52. Helander A, Holttinen H, Paatero J (2010) Impact of wind power on the power system imbalances in Finland. Accepted to *IET Renewable Power Generation Journal*
 53. FGE/FGH/ISSET: Bewertung der Optimierungspotenziale zur Integration der Stromerzeugung aus Windenergie in das Übertragungsnetz (2007) <http://www.erneuerbare-energien.de/inhalt/42024/4591/>
 54. Holttinen H, Koreneff G (2007) Imbalance costs of wind power for a hydro power producer in Finland. In: Proceedings. European Wind Energy Conference EWEC2007. Milan, Italy, 7–10 May, 2007. European Wind Energy Association, EWEA
 55. Energinet.dk: System Plan (2007) www.energinet.dk
 56. Eriksen PB, Orth A (2008) Challenges and Solutions of Increasing from 20 to 50 Percent of Wind Energy Coverage in the Danish Power System until 2025; Invited Keynote Paper. In: Proceedings of the 7th international Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms; 26–28 May 2008, Madrid, Spain
 57. Ummels BC (2009) Power system operation with large-scale wind power in liberalised environments. PhD thesis, Delft University of Technology, The Netherlands, 192 p
 58. All Island Grid Study (2008) <http://www.dcenr.gov.ie/Energy/North-South+Co-operation+in+the+Energy+Sector/All+Island+Electricity+Grid+Study.htm>
 59. The Effects of Integrating Wind Power on Transmission System Planning, Reliability, and Operations. Report on Phase 2, GE Energy. Prepared for The New York State Energy Research and Development Authority, City, State, March 2005. http://www.nyiso.com/public/services/planning/special_studies.jsp
 60. Van Hulle F, Tande JO, Uhlen K, Warland L, Korpås M, Meibom P, Sørensen P, Morthorst PE, Cutululis N, Larsen H, Woyte A, Verheij F, Kleinschmidt C, Moldovan N, Holttinen H, Lemström B, Uski-Joutsenvuo S, Gardner P, Purchala K, Tiedemann A, Kreutzkamp P (2009) Final Report TradeWind. <http://www.trade-wind.eu>
 61. Matevosyan J (Riga Technical Univ., Riga, Latvia) (2004) Estimation of potential wind energy curtailment for wind power plants in power systems with bottleneck problems; Soder L (2004) EPE-PEMC 2004 11th International Power Electronics and Motion Control Conference, vol 5, pp 5–502

Books and Reviews

- Söder L (2004) On limits for wind power generation. *Int J Global Energy Issues* 21(3):243–254
- Söder L (2005) The value of wind power. In: Ackermann T (ed) *Wind power in power systems*. Wiley, ISBN 0-470-85508-8

Wind Power Generator Systems and Local Power System Interconnection

TORBJÖRN THIRINGER

Department of Energy and Environment, Division of Electric Power Engineering, Chalmers University of Technology, Göteborg, Sweden

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Operational Requirements on a Wind Turbine
 Generator System: General
 Wind Turbine System Electrical Components
 Power Electronics
 Wind Turbine Generating Systems
 Power Quality Impact

Fixed-Speed Wind Turbines
 DFIG and Full Converter Turbines
 Behavior During Network Faults
 Future Directions
 Abbreviations
 List of Commonly Used Variables
 Bibliography

Glossary

Converter Electronic object that can change a voltage and current form into another form.

Doubly fed induction generator system (DFIG) A common electrical system for wind turbines that has a converter in the rotor circuit.

Fault-Ride-Through – FRT The ability of a wind turbine to be immune against voltage dips in the connected network.

Fixed-speed system Simplest wind turbine electrical system consisting of mainly an induction machine and capacitors.

Full-power systems Wind turbine electrical system which has a converter between the stator of the wind turbine generator and the grid transmitting the whole power from the turbine.

Power quality An expression for how close the voltage is to its ideal waveform, and the quality of the current injected by an object into the grid.

Reactive power Oscillating power that only causes losses and do not provide an active power transfer.

Definition of the Subject

The electrical system of a wind turbine could be referred to as the heart of the wind turbine. The main purpose of the electrical system is of course to convert the incoming mechanical energy on the shaft to electrical energy that is fed into the connected power grid. This conversion should naturally be done with a minimum of losses as well as with the smallest component size possible.

In addition, the electrical system has the ability to influence the power quality impact that a wind turbine has on the power grid and also the ability to control the speed and torque on the machine shaft (if it is equipped with a power electronic converter) and in this way maximize the energy capture as well as to keep the stresses on the structure to a minimum.

These requirements (high energy capture, good power quality impact, low stresses on the structure in order to minimize the material needed) have in combination with the technical development led to an evolution of wind turbines and definitely regarding the electrical systems of wind turbines. This development is continuously ongoing and is the reason to why we have several different concepts of electrical generating systems for wind turbines.

This entry explains the commonly used technologies for generating systems for wind turbines today, their design, function, and properties. In addition, it also presents how turbines made up of the various technologies influence the grid to which they are attached.

Introduction

Wind turbines, or rather wind mills, have existed for millenniums, mainly for crop grinding and water pumping purposes. In 1887 the first electricity-generating wind turbine was erected and used for battery charging. During the years the sizes of the wind turbines increased and became more and more connected to the grid for conventional power infeed generation. 1931 a 100 kW wind turbine was erected in the USSR. The first megawatt (MW) machine was erected in the USA in 1941 [1].

However, due to the enormous amounts of available fossil fuels, the more expensive wind energy electricity was no greater success during the greater part of the twentieth century. However, with the oil crises in the beginning of the 1970s the interest awoke and the modern era of wind energy started. In many countries, for instance in Denmark and in the USA, more modern wind turbines started to be built and erected. By the beginning of the 1980s, the standard maximum size was around 50 kW and at the end of the 1980s 200 kW. Ten years later, the maximum size was up to around 2 MW and 10 years further 6 MW turbines existed. The most economical size has typically been a bit smaller than the maximum size. Another important shift happened in the 1990s, the variable-speed operation was introduced into commercial turbines after having mainly been a research topic.

In short, the development of electrical systems for wind turbines has moved from being small isolated grid disturber to being large grid-friendly and even

grid-reinforcing production units. This development has to a great extent been built on the development of power electronics. Today (2011) there is a lot of focus on large wind farms connected to the transmission networks, but local integration of wind turbines will also continue to be an important issue.

Operational Requirements on a Wind Turbine Generator System: General

Stationary Requirements

As partly mentioned in the previous section, a wind turbine generator system shall have a low weight, low maintenance need, as well as a high efficiency. The resulting goal is of course, that it shall be as cheap as possible in relation to the produced electrical energy. It shall also be able to convert the maximum power without becoming overheated. The sizing of the electrical system of course follows the maximum rating of the wind turbine, in which it is going to be inserted, and the electrical system must be able to handle this power as well as some overload. The efficiency should preferably not be the best for the rated operation, instead it is more important that the efficiency is as highest for medium speeds. In Fig. 1 the wind distribution for a site with an average wind speed of 7 m/s is presented, as well as the available energy for each specific wind speed, which demonstrates the focus area for having a high efficiency of the electrical system.

From Fig. 1 it is very clear that high efficiency around 1–2 times the mean wind speed is a key issue for the electric generating system. Moreover, it can be

noted that below 3 m/s and above 20 m/s there is hardly any energy at all.

The system must have the ability to be overloaded for short times, without losing its functionality. Electrical machines have in this regard a highly desirable property, they can be overloaded for a short time, since the magnetic ability of the generators are usually much higher than the steady-state thermal limitation. However, this is not the case with the power electronic converters; they usually cannot even handle a few seconds of higher power compared to their rating.

Dynamic Requirements

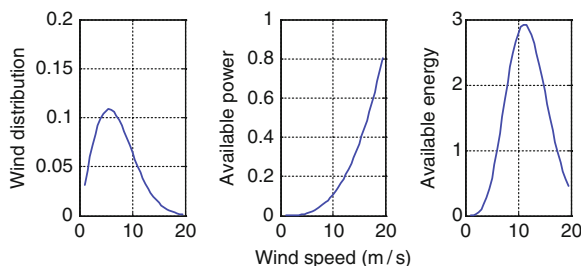
In this section, there appear a number of terms which will be declared in the upcoming section. The main purpose of this section is to underline the great benefit of having an electrical generating system that can control the speed and torque of a wind turbine in a broad range.

Apart from the pure stationary demands on proper rating and high efficiency, a very important issue is that the electrical system of the wind turbine must provide a sufficient smoothness regarding the electric power flow in order not to cause stresses either on the structure or on the power quality of the grid.

For a traditional fixed-speed wind turbine, there is a trade-off between softness and efficiency. It is the rotor resistance in the induction generator used in this system that is the key variable. This together with the input torque determines that a fixed-speed system should be stall controlled. The key equation here is

$$J \frac{d\omega}{dt} = T_w - T_e$$

where J is the inertia of the system, T_e is the electrodynamic torque, and T_w is the torque from the wind acting on the rotor. In the case that the rotor speed cannot change, the electrodynamic torque will be equal to the shaft torque created through the aerodynamic power conversion by the rotor, which means that all unevenness in the incoming torque is transmitted further in the system. This causes both mechanical wear and a poor power quality impact. However, for fixed-speed systems, the speed can change a little bit, and even more if the induction generator has a high rotor resistance, which unfortunately gives a lower efficiency. Anyway, the speed variations for a



Wind Power Generator Systems and Local Power System Interconnection. Figure 1

Using 7 m/s as average wind speed. Wind speed distribution, available power in the wind and available energy as a function of wind speed

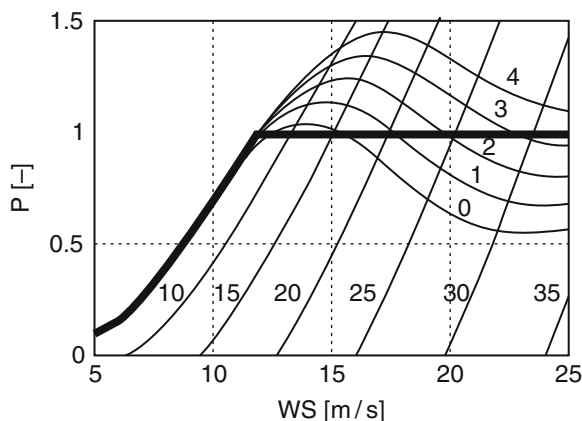
fixed-speed wind turbine system are low. In Fig. 2 the power versus wind speed for some different pitch angles are presented. Pitch angle is a quantity that tells how “turned” the blade is toward the rotational plane.

From the figure it can be noted that a wind increase around 20 m/s will lead to a very large power variation if a pitch angle of 20° is used, which will lead to stresses on the mechanical construction as well as a power output that varies substantially. Accordingly stall control must be used for a fixed-speed turbine, where the power variations are much lower. This can be noted by observing the 2° and 25° pitch angle curves at 20 m/s, where the turbine is producing rated power. An increase in 2 m/s, leads to a reduction of the power by 10% for the 2° angle while it for a 25° angle leads to an increase of 50%, which would strongly stress the turbine.

Another observation from the figure is that it can be noted that a fixed pitch angle at 0° would be the safe choice for this turbine. A drawback is then that some energy is lost in the high wind speed region. If some overpower can be accepted a slightly higher pitch angle could be possible to use. Important to point out is also that with changing air pressure, the power production will vary.

Power Versus Wind Speed Relation

The equations used to obtain the static power versus wind speed presented in Figs. 1 and 2 can be easily



Wind Power Generator Systems and Local Power System Interconnection. Figure 2

Wind speed versus power (1 corresponds to full power) for various pitch angles for one given pressure

found in literature, for instance in [2], and the key equations for doing these calculations are

$$P_{\text{wind}} = \frac{1}{2} \rho A W S^3$$

$$P_{\text{mech}} = \frac{1}{2} \rho A W S^3 C_p(\lambda)$$

$$\lambda = \frac{\omega_r R}{W S}$$

where P_{wind} is the available power in the wind, P_{mech} is the power that can be converted from P_{wind} , R is the rotor diameter, A is the area swept by the rotor, ρ is the air density, λ is the tip-speed ratio, C_p is the power coefficient determining the ratio between P_{mech} and P_{wind} .

C_p is a function of λ . For a specific λ , C_p has it highest value, about 0.45–0.5. This means that if the rotor speed can vary proportionally with the wind speed, the power capture can be maximized. In order to vary the rotor speed, a variable-speed turbine is needed. For higher wind speeds, the available power in the wind becomes far too high, so C_p must go down in order to limit the power. This reduction can be done in two ways, either the blades lose their aerodynamic efficiency, “stall control,” or the blades are turned out of the wind, “pitch control.”

If the rotor speed was allowed to vary, as for a variable-speed turbine, it can be noted from the earlier presented speed equation that there is a term, $J d\omega/dt$, that can take up the incoming power variations and give a smooth power output (Power is directly related to T_e). By selecting a proper power electronic control, the electrodynamical torque can be controlled and then also the stresses on the mechanical structure as well as and the power fluctuations going out to the grid, can be reduced. For a variable-speed turbine, pitch control is generally used, since pitch control gives less stresses on the wind turbine. In this case, the turbine can speed up in case of a gust (keeping the shaft torque constant through the generator system) and during the acceleration, the additional energy is stored in the rotational energy and then the turbine can pitch the blades to return to a suitable aerodynamic power conversion. This gives a requirement on the torque control ability of the generating system; it must have a bandwidth of at least 10 Hz. This is no

problem for electrical generating systems in general, where torque demands can be realized in a matter of less than 10 ms.

In addition, the energy capture by the rotor can be increased since an optimal speed can be used for each wind speed. However, this is a cost issue. Today, the fixed speed systems often comprises two fixed speeds, and in this way, modern fixed-speed systems have about the same energy production as variable speed systems.

Wind Turbine System Electrical Components

Electrical Fundamentals

Voltage, Currents, Power, Energy, and Loss In electrotechnology, the intensity source is voltage and resulting flow is current. The units are Volts [V] for the voltage and Ampere [A] for current. There are three basic electrical properties related to passive elements in electrotechnology, namely, the resistance, the inductance, and the capacitance. The relation between currents and voltages are for these three elements:

$$u(t) = Ri(t)$$

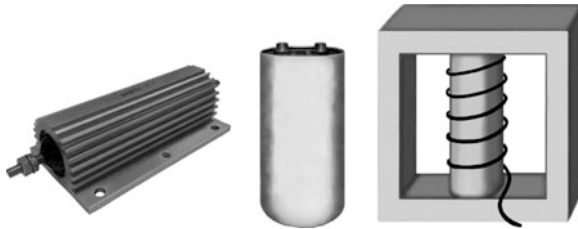
$$u(t) = L \frac{di(t)}{dt} \text{ or } i(t) = i(t_0) + \frac{1}{L} \int_{t_0}^t u(t) dt$$

$$i(t) = C \frac{du(t)}{dt} \text{ or } u(t) = u(t_0) + \frac{1}{C} \int_{t_0}^t i(t) dt$$

In Fig. 3 below, a resistor, a capacitor, and an inductor are shown.

The power flowing in an electrical system is

$$p(t) = u(t)i(t).$$



Wind Power Generator Systems and Local Power System Interconnection. Figure 3

Resistor, capacitor, and inductor

By integrating the instantaneous power, the energy can be determined as

$$W(t) = \int_0^t p(t) dt.$$

The losses in electrical systems are typically related to resistances in the systems and can directly be found as

$$p_{\text{loss}}(t) = Ri^2(t)$$

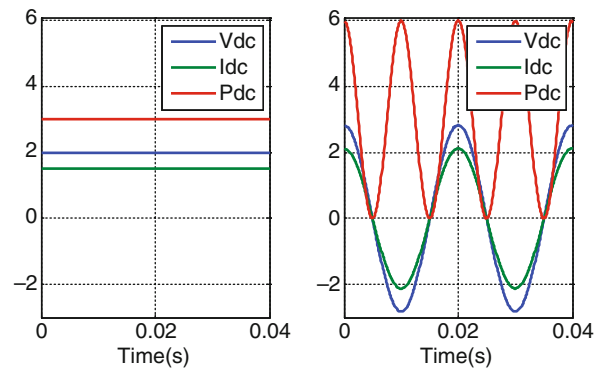
or

$$p_{\text{loss}}(t) = \frac{u^2(t)}{R}.$$

where i is the current flowing through the resistive element and u is the voltage over it.

It is possible to use constant voltage and current levels as well as an alternating voltage and current in an electrical system; these are referred to as Direct Current = DC and Alternating Current = AC. HVDC accordingly means High Voltage Direct Current. The frequency of the alterations is typically 50 or 60 Hz in grids around the world. In Fig. 4 below both DC-quantities as well as AC-quantities (50 Hz) are depicted.

This type of AC-supply is referred to a single-phase supply, more about the commonly used three-phase supply will soon follow. Here some observations can be performed. The average power in both cases in Fig. 4 are 3 W. In the case of the AC-system the instantaneous power however oscillates with twice the frequency of the voltage and current, which is a drawback when using



Wind Power Generator Systems and Local Power System Interconnection. Figure 4

Left: DC-quantities. Right: AC-quantities

single-phase AC-system for power electronic equipments, where this power variation needs to be tackled.

Moreover it can be observed that the peak values of the currents and voltages are higher than the quantities in the DC-system. Due to the constant changing of the level of current and voltage, it is in an AC-system necessary to have a peak value that is $\sqrt{2}$ higher than the so-called rms-value:

$$\hat{U} = \sqrt{2}U_{\text{RMS}}$$

The RMS-value approximately/exactly corresponds to the DC-level, in this example exactly. For further information, the reader is encouraged to look in a textbook on the subject, for instance [3].

In the case treated in Fig. 4, the voltage and currents were exactly in phase. However, when applying AC-voltage on a circuit consisting of inductances and capacitances, there will be a phase shift. In Fig. 5 an R-L circuit subjected to an alternating voltage, v , and the resulting quantities are displayed.

One obvious observation is that the current now is lagging the voltage and that the power level is lowered, although the peak voltage and peak current are the same. This gives rise to three different powers.

The active and reactive power is defined as

$$P = UI \cos(\phi)$$

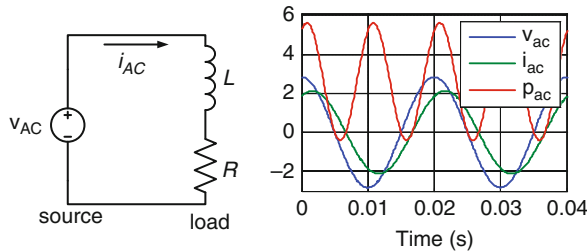
and

$$Q = UI \sin(\phi).$$

The active and reactive power are related as

$$S = \sqrt{P^2 + Q^2}$$

where S is the apparent power.



Wind Power Generator Systems and Local Power System Interconnection. Figure 5

A voltage source acting on an R-L load

It is the apparent power that is important for the electrical dimensioning of a component since it composes the peak value of the current and voltages. The active power is the desired one and the reactive power is usually undesired. An important quantity that tells how much of the transmitted power that is useful is the power factor:

$$PF = \cos(\phi)$$

These four equations above are only valid in the absence of harmonics. In the presence of harmonic voltages, there is a need for more definitions. Here the reader is encouraged to consult power electronic literature, for instance, [3].

Three-Phase System It is possible to use three single-phase systems where the AC-voltages are separated with one-third period apart to form a three-phase system. This corresponds to an angle displacement of 120° between each of the single-phase systems.

$$u_R(t) = \sqrt{2}U_{\text{RMS}} \sin(\omega t)$$

$$u_S(t) = \sqrt{2}U_{\text{RMS}} \sin(\omega t + \frac{2\pi}{3})$$

$$u_T(t) = \sqrt{2}U_{\text{RMS}} \sin(\omega t + \frac{4\pi}{3})$$

$$i_R(t) = \sqrt{2}U_{\text{RMS}} \sin(\omega t + \phi)$$

$$i_S(t) = \sqrt{2}U_{\text{RMS}} \sin(\omega t + \frac{2\pi}{3} + \phi)$$

$$i_T(t) = \sqrt{2}U_{\text{RMS}} \sin(\omega t + \frac{4\pi}{3} + \phi)$$

In Fig. 6, three-phase currents and voltages are shown for a case with $U_{\text{RMS}} = 230\text{V}$, $I_{\text{RMS}} = 150\text{A}$ and $\phi = 30^\circ$.

An advantage is that the instantaneous power

$$p(t) = u_R(t)i_R(t) + u_S(t)i_S(t) + u_T(t)i_T(t)$$

now is continuously constant, so that in power electronic systems using a three-phase supply, there is less need of filtering capacitors in order to obtain stiff voltages after the in power electronic equipment very common rectification. Moreover,

$$i_R(t) + i_S(t) + i_T(t) = 0$$

and

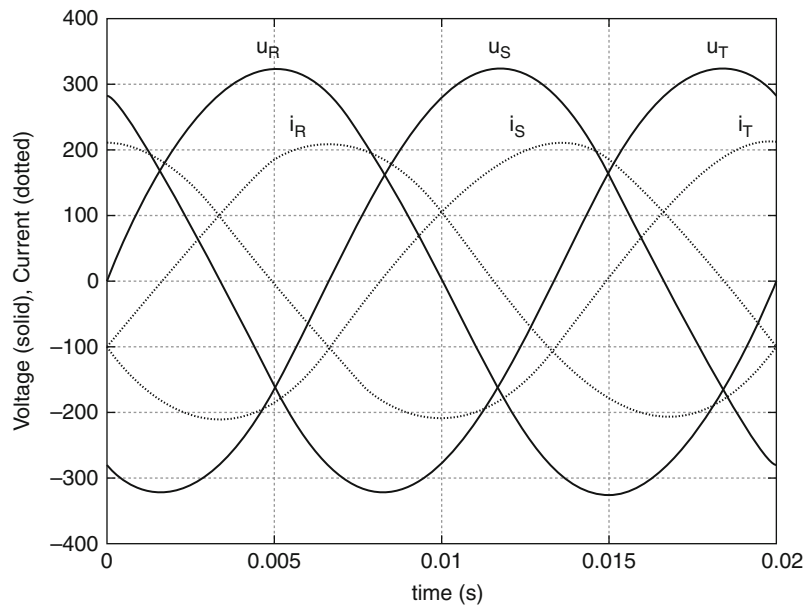
$$u_R(t) + u_S(t) + u_T(t) = 0$$

which means that there is no need for a return conductor, so more power can be transmitted per conductor. This is illustrated in Fig. 7.

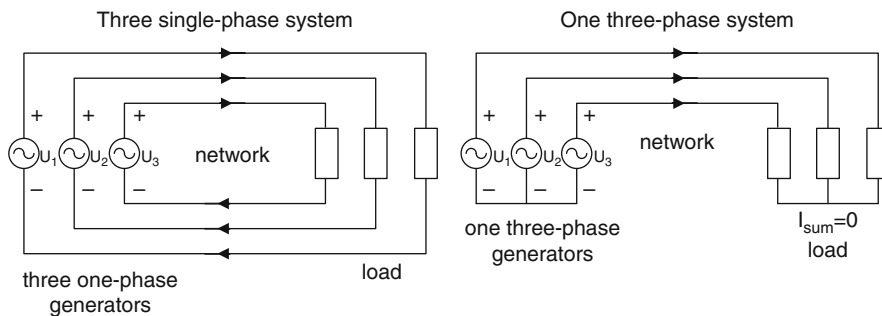
These nice properties together with the invention of the three-phase induction machine (see coming section) have led to a condition where three-phase systems completely dominate the electric power transmission as well as power supply today.

Jω-Method A convenient mathematical method that can be used to calculate the currents and voltages if the quantities are in steady state and oscillating with one frequency only, is the *jω*-method, [2]. The sinusoidally varying quantity is represented with a complex number where the magnitude and phase shift are represented. The representation can be done either in Cartesian coordinates or in polar coordinates.

$$A + jB = \sqrt{A^2 + B^2} e^{j \arctan(\frac{B}{A})}$$



Wind Power Generator Systems and Local Power System Interconnection. Figure 6
Three-phase currents and voltages



Wind Power Generator Systems and Local Power System Interconnection. Figure 7
Three single-phase systems as well as a three-phase system

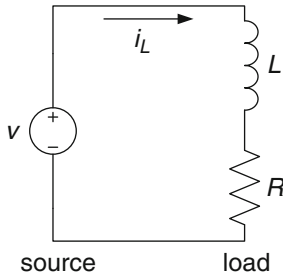
When a polar coordinate system is used, the magnitude in combination with the angle is referred to as a phasor.

Due to the steady-state assumption and a single oscillating frequency, the derivative d/dt can be replaced with $j\omega$, and then calculations can be performed, see the example when determining the currents in an RL-circuit. In Fig. 8, an example of an RL-circuit supplied by an alternating voltage source is shown.

The equation for the circuit is

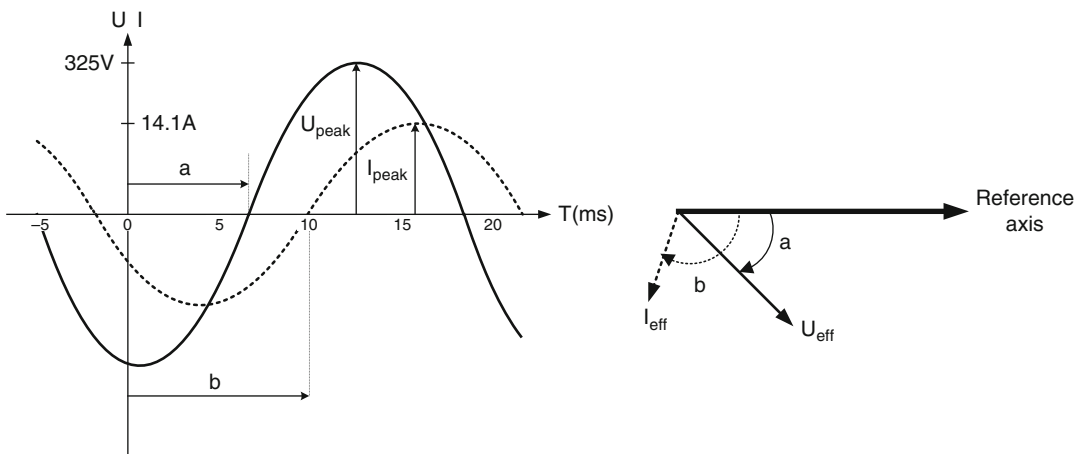
$$u(t) = Ri(t) + L \frac{di(t)}{dt}.$$

or using the $j\omega$ -method,



Wind Power Generator Systems and Local Power System Interconnection. Figure 8

A voltage source acting on an RL load



Wind Power Generator Systems and Local Power System Interconnection. Figure 9

Voltages and currents as time traces as well as in phasor representations

$$\underline{U} = R\underline{I} + j\omega L\underline{I}.$$

The resulting waveforms are shown in Fig. 9.

The Transformer Power is voltage times current. In order to have a convenient end-usage a lower voltage level is needed, while for transportation a higher voltage is needed to keep the losses, that is, the currents down.

Due to the simplicity with which voltage could be changed into different voltage levels, alternating-current systems (AC-systems) became the dominating ones for higher power already around year 1900. The object that transforms voltage from one voltage level to another one is called a transformer, see Fig. 10 where two photos of a large as well as a small transformer are shown together with the symbol of a three-phase transformer.

The equations governing the voltage and current relations (Observe: a bit idealized) between the primary and secondary side are the following:

$$\frac{U_1}{N_1} = \frac{U_2}{N_2}$$

$$I_1 N_1 = I_2 N_2$$

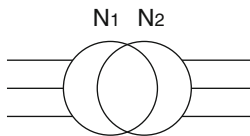
Electric Machines (Generators and Motors)

Electric machines can generally both convert rotational energy into electrical energy (generator operation) and the opposite (motor operation). When an electric

machine is designated to produce electrical energy from another source, the term “generator” is typically used. Here the more general term “machine” will frequently be used.

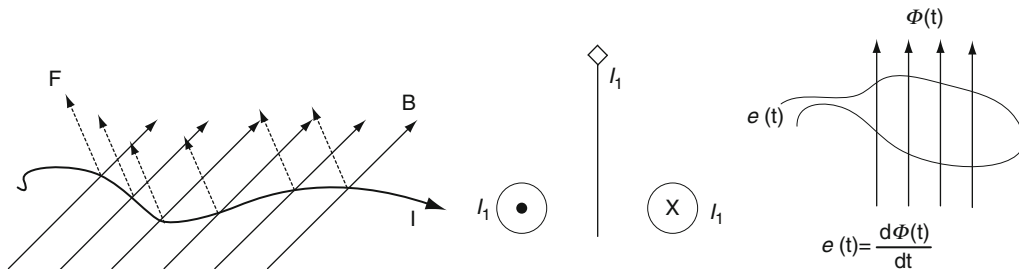
There are three main types of electrical machines. These are the DC-machine, the induction machine (asynchronous machine), and the synchronous machine. They all consist of a rotor that rotates, and usually is located in the inside of the machine and a stator, which is standing still.

There are three fundamental principles that are used in an electrical machine, which gives rise to a conversion between electrical and mechanical energy.



Wind Power Generator Systems and Local Power System Interconnection. Figure 10

Transformer, photo and circuit description, N_1 is the number of windings on the primary side and N_2 is the number of windings on the secondary side



Wind Power Generator Systems and Local Power System Interconnection. Figure 11

Three electromagnetic fundamentals making up key factors for the operation of an electrical machine. Same order as the bullets above. See [4] for more details

- A force is produced by the cross product of a current flowing in magnetic field $F = B \times I$. If this force is applied at a certain distance, r , from the axis of the rotor, a torque will be formed according to $T = F \times r$. Force can also be looked upon as being created by two interacting magnetic fields, more about this follows below.
- A magnetic field is created by a permanent magnet or an electromagnet.
- A flux change through a coil leads to an induced voltage in this coil. $e = \frac{d\psi}{dt}$

Figure 11 illustrates in a simple fashion the three physical fundamentals, for a more thorough description, [4] is recommended.

There is an alternative way of looking into the force creation, which is illustrated in Fig. 12.

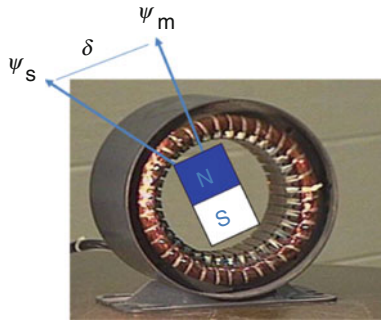
By having a magnetic vector in the rotor and one in the stator, the rotor will be exposed to a force

$$T = |\psi_s||\psi_r| \sin(\delta)$$

and the result will be that the rotor would like to align itself with the magnetic vector of the stator. If the magnetic vector (flux vector) now rotates, the rotor will also rotate. Some recommended textbooks for further information are [4–7].

DC-Machine Although the primary interest here is to use the machines as generators, the more general term “DC-machine” will frequently be used in this section.

In order to create the varying flux through a coil, either the coil must move (as in a DC-machine) or the flux through the coil must vary (AC-machine).



Wind Power Generator Systems and Local Power System Interconnection. Figure 12

Illustration of one magnetic vector in the rotor and one imposed in the stator giving an alignment force, making the machine to rotate

In the DC-machine, this “fluxtaking coil” is located in the rotor and accordingly there is a need for a mechanical shifting of the current. This is done in something called the commutator. Due to practical reasons, such as that the machine becomes expensive, the commutator has reliability problems and that the power level is close to the maximum for this machine, the DC-machine is nowadays hardly not used for wind energy applications, and in this case limited to very small turbines.

In Fig. 13 the stator and the rotor of a DC-machine is presented. The field winding creates the field, and when the rotor is inserted and the rotor conductors are carrying a current in the magnetic field originating from the field winding, forces are formed which will create a rotation of the DC-machine.

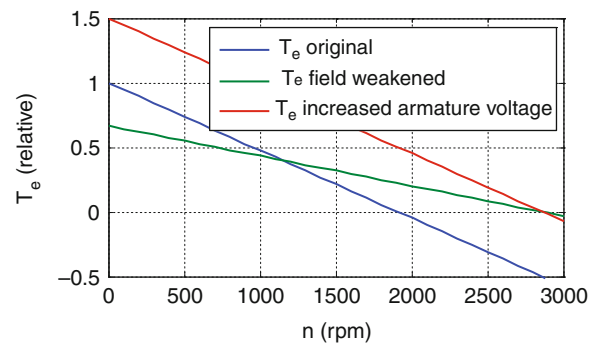
The separately magnetized DC-machine has the torque–speed characteristic presented in Fig. 14. The torque presented is the electrodynamic torque, T_e , which is the torque that the machine sets up based on the electrical conditions. The blue curve corresponds to an original setting. By adjusting the field voltage or the armature voltage, it is possible to alter the torque–speed characteristic and in this way obtain a torque or speed control. Earlier, before converters for AC-machines became useful, it was the DC-machine that was used for variable-speed drive applications.

In the figure, a negative torque corresponds to generator operation. For instance, note the blue curve for speeds above 1,900 rpm. This means that if a driving wind torque is applied, the turbine will speed up



Wind Power Generator Systems and Local Power System Interconnection. Figure 13

Left: Stator with field winding. *Right:* Rotor with armature winding



Wind Power Generator Systems and Local Power System Interconnection. Figure 14

Torque–speed characteristics of the separately magnetized DC-machine

until the incoming torque is balanced by the electrodynamic torque set up by the DC-machine.

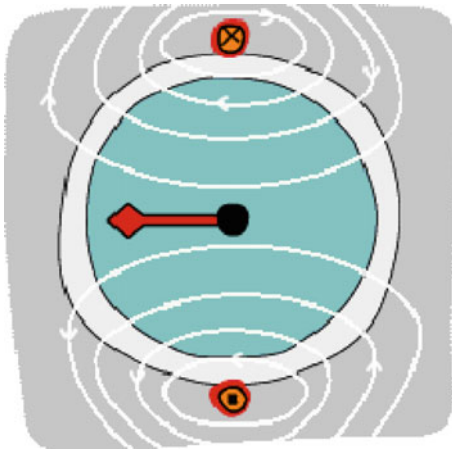
Rotating Flux Vector Before continuing with the alternating current machines, there is a need to tackle the issue that an AC-machine will not have a “room-fixed” flux, instead it will rotate. Accordingly, this section tries to explain how a rotating flux vector is built up from the voltages applied to the machine and from the resulting currents.

By feeding a coil with an AC-voltage, an AC-current will result and a pulsating magnetic flux will be created. The direction of the magnetic flux follows the magnetic axis of the coil which is perpendicular toward the coil. The direction can be determined using the right hand, with the fingers in the direction of the current in the coil, and then the magnetic flux direction is in

the direction of the thumb. In Fig. 15, the magnetic vector is depicted for one time instant.

If several coils are acting on the same magnetic path, there will be a total magnetic vector composed of all coils that act on this magnetic path. By using three coils that are displaced 120° “roomwise” and feeding them with voltages displaced 120° in angle (time), a rotating magnetic vector will be created, see illustration in Fig. 16. In the figure, three time instants are presented, where it can be noted how the gray arrow representing the total magnetic vector rotates with a constant speed and constant magnitude.

Depending on the rotor design, a flux vector will be created in the rotor and the angle between the applied



Wind Power Generator Systems and Local Power System Interconnection. Figure 15

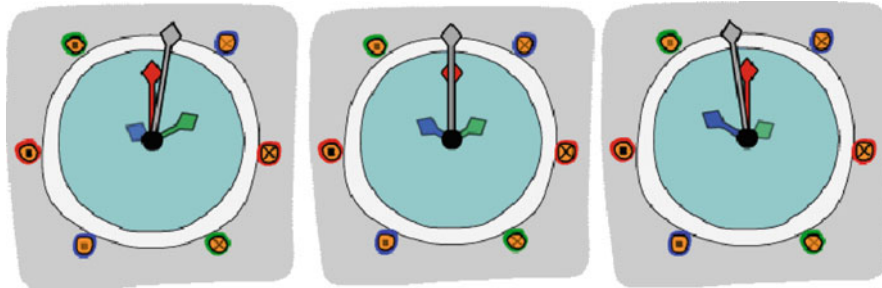
Magnetic flux created by current flowing in a single-phase coil

magnetic vector from the stator and the rotor flux vector will give rise to a driving torque and a rotation that has the same speed as the rotating magnetic flux vector originating from the stator (synchronous machine), or slightly lower (induction machine).

By placing the magnetic vector in the stator before the magnetic vector from the rotor, the machine will operate in motor operation, and by putting the magnetic vector from the stator behind the magnetic flux from the rotor (and thus braking the rotor) the machine will operate as a generator.

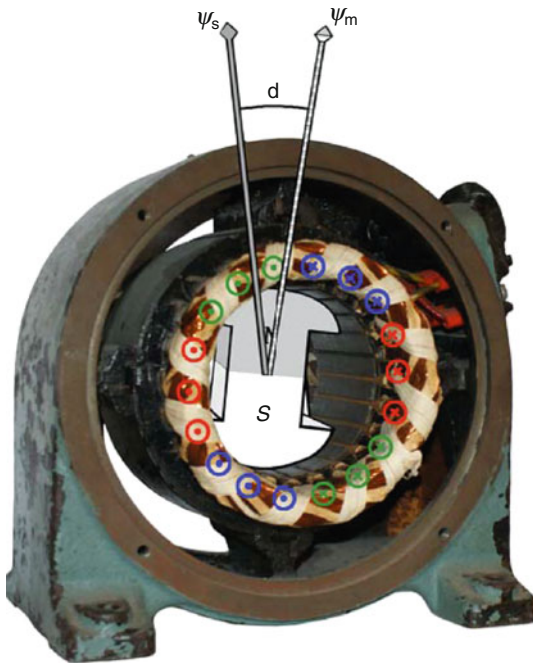
Synchronous Machines The synchronous machine has its name since the rotor will rotate synchronously with the stator flux vector. The machine can have (1) permanent magnets, (2) an electrical winding, or (3) just a magnetically uneven rotor shape. In Fig. 17, a simplified figure of a permanent-magnet synchronous machine is presented.

As earlier mentioned, a machine with a magnetic vector created in the rotor by permanent magnets or a field winding are made to rotate as motors by applying a rotating magnetic vector created from the stator, before the magnetic vector of the rotor. In this way, the stator magnetic vector is “pulling” the rotor behind itself. In generator operation, it is the load (wind/water/steam) that makes the rotor to rotate, and then, in order to brake the machine so it does not overspeed, a rotating magnetic vector created by the stator, currents must exist. Usually this magnetic vector is created by applying voltages to the stator, but it is also possible to allow voltages to be generated in the stator by the field vector from the rotating rotor, and then



Wind Power Generator Systems and Local Power System Interconnection. Figure 16

Magnetic flux created by a three-phase coil, forcing the machine to rotate counterclockwise



Wind Power Generator Systems and Local Power System Interconnection. Figure 17

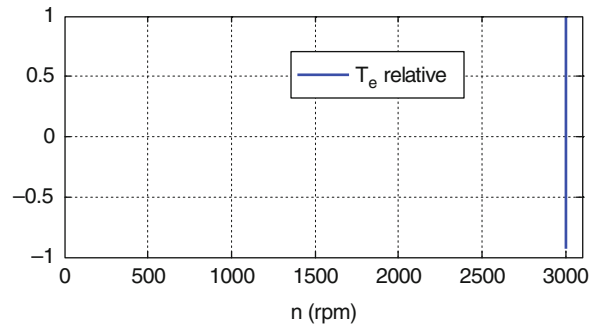
Simplified picture of permanent magnet machine, with stator flux (gray arrow) and magnet flux (gray-white arrow).

connect the stator windings to a passive load (a load that can be represented by an RLC-circuit).

The torque–speed characteristics for a two-pole synchronous machine is presented in Fig. 18 for a frequency of 50 Hz. If the rotor has more poles, it will rotate slower; a four-pole machine would with 50 Hz rotate with 1,500 rpm. If a voltage having a 60 Hz frequency is applied, the machine would rotate a bit quicker; more about this will come below in the pole-pair section.

Induction (Asynchronous) Machines A photo of an opened three-phase induction machine is presented in Fig. 19 together with a rotor.

From the figure, the stator windings can be observed, located around the rotor with their magnetic axes directed in to the rotor. By using the right hand, and pointing the thumb inward from the stator in to the rotor, the other fingers on the right hand, depicts how the windings in the stator are placed with the fingers showing the current direction. Larger machines



Wind Power Generator Systems and Local Power System Interconnection. Figure 18

Speed–torque characteristics of a synchronous machine



Wind Power Generator Systems and Local Power System Interconnection. Figure 19

Photo of an “opened” three-phase induction machine as well as a rotor originating from another machine

are in principle almost always three-phase, and accordingly, as described above, there are three magnetic axes pointing inward, displaced 120° from each other around the rotor.

Further, the rotor can also be observed in the photos. The rotor consists of laminated iron sheets/plates, which can be assumed to be isolating toward currents. In these laminations, however, there are cast aluminum slots, where a current can flow. If the stator of the induction machine is supplied by a three-phase power source, a rotating magnetic flux will be created. As this magnetic wave rotates through the rotor, the short-circuited current paths (that can be observed in the figure) try to put up currents that counteract this imposed magnetic field. These currents form together with the imposed flux vector a driving torque. The process with which the currents are created in the rotor is called induction, and hence the name induction motor. If the machine rotates exactly with the speed of the magnetic flux vector set up by the stator,

there will be no induced currents and the torque will be zero. However, the larger the differences between the speed of the rotor and the speed of the magnetic field, the larger the torque will be. However, a too large speed difference will cause a reduction of the torque and the machine will lose its thought functionality. The point where the torque is at its maximum is referred to as the stall point. In Fig. 20, the flux vector (white) together with the current vector in the rotor (gray) is presented for a lightly loaded motor point as well as for operation at full speed as a generator.

In the figure it can be noted that, exactly as for the DC-machine that the speed increases in generator operation. If the induction machine is driven by a wind turbine rotor, it is the torque formed from the wind that increases the speed.

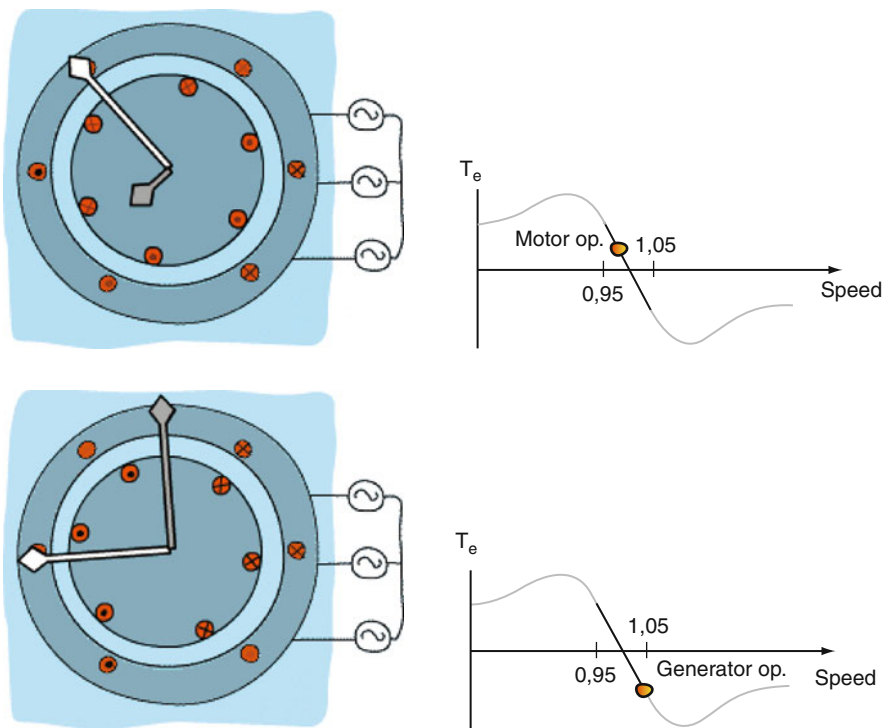
The fact that the induction machine operates asynchronously with the frequency of the supply gives it its alternative name, the asynchronous machine.

The machine presented in Fig. 20 has a rated slip of 5% (slip is the relative speed change that the machine

has within its normal operating region). Larger machines have a rated slip of below 1%. In [8], page 124, a figure presents typical slip values for induction generators as a function of size and pole-pair number. The slip increases with higher rotor resistance, which gives a smoother operation, since the speed is allowed to vary more. However, the higher the rotor resistance, the higher the losses. The efficiency of a modern induction generator for wind turbine applications is above 95% once the loading is between one-third of full load and full load.

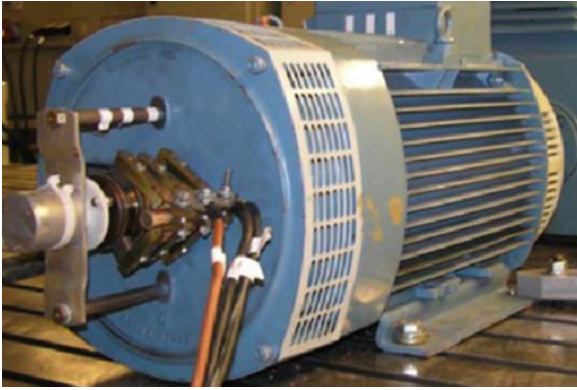
Of special interest for wind energy applications is an induction machine where the rotor instead has windings and these windings are connected to slip rings at the end of the machine. In Fig. 21, slip rings that are connected to the rotor windings can be observed on a 22 kW induction machine.

To the rotor slip rings it is possible to connect a converter. The advantage is that such a converter becomes fairly small in comparison with attaching a converter in the “normal way,” that is, to the stator,



Wind Power Generator Systems and Local Power System Interconnection. Figure 20

Upper picture, operation of the induction machine as a motor with a moderate torque. Lower picture, operation as a generator with full torque. White vector is the resulting magnetic vector, and the current vector is depicted in gray



Wind Power Generator Systems and Local Power System Interconnection. Figure 21

Slip rings connected to the rotor windings of a 22 kW induction machine

giving low cost and low losses in the generating system of the wind turbine since only a fraction of the total produced power goes through the converter. More about this will follow in the descriptions of the wind energy generating systems.

Pole-Pair Number If there only is a north and a south pole in the stator and the rotor, the synchronous machine will rotate with the same frequency as the applied electrical frequency

$$\Omega_m = \omega_{el}.$$

The induction (asynchronous) machine will, as mentioned before, deviate maximum a few percent from the speed determined by the equation above.

The rotational speed in rpm (round per minute) can be found using the relation

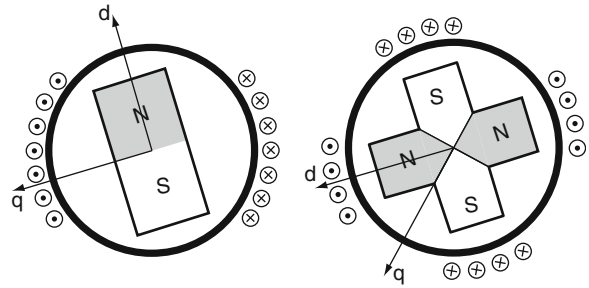
$$n = \Omega_m \frac{60}{2\pi}.$$

Often, an electrical machine consists of more than a south pole and a north pole (referred to as 1 pole-pair, with the notation $p = 1$). In Fig. 22, an example of a two- and a four-pole machine is depicted.

The relation now becomes for the synchronous machine,

$$p\Omega_m = \omega_{el}.$$

Here, the magnetic flux per pole-pair changes with the same speed as the applied frequency. However,



Wind Power Generator Systems and Local Power System Interconnection. Figure 22

Two- and four-pole synchronous machine

since there are several pole-pairs involved, this means that the speed of the rotating flux vector is shifted down with the number of pole-pairs, seen from the stationary reference system. Apart from giving the possibility to change the rotational speed, the use of a higher number of pole-pairs can also give a possibility to optimize the machine for certain applications.

Again, for the induction machine there is a slight speed difference between the physical rotor speed and the speed of the magnetic vector, but the same reasoning holds for a no-loaded induction machine.

Power Electronics

Converter

The rotational speed of an AC-machine is governed by the number of pole-pairs as well as the frequency of the electric supply it is connected to. This is the reason why traditional (also called Danish concept) turbines are called fixed-speed turbines. This type of machine is put in generator operation by the driving wind, but the speed is governed by the grid frequency 50/60 Hz, the number of pole-pairs, and the gearbox ratio.

In order to control the speed of an AC-machine, it is a necessity to be able to alter the frequency of the voltage applied to the electric machine. In this way, the magnetic flux vector will alter its speed. For this purpose a converter is used. The converter has an AC-side connected toward the generator and a DC-side that sends the electrical power toward the grid, via another converter. This second converter has its AC-side toward the grid.

The needed varying frequency is created by switching transistors quickly (a couple of kHz) and in

this way a switched voltage pattern is formed, that in average (for a switching period, i.e., 0.5–1 ms) has a desired average that has a “main fundamental waveform” that corresponds to the desired applied frequency. In Fig. 23, a simplified sketch of a converter is depicted together with a 200 A power module making up a so-called phase leg, which builds up the phase voltages, in this case phase *c*. The transistor is either on or off, and in a phase leg, the upper or lower is always on and the other off. This means that the voltage potential in point *c* jumps between $+V_d/2$ or $-V_d/2$. Once the voltages are created by the transistors, the current will then flow out from the generator and towards the DC-link. Also in the figure, bottom right, a complete conversion from the variable generator frequency to the 50/60 Hz grid frequency is depicted.

Since there are three-phase voltages created, the phase potentials will contain also another level. The resulting phase voltage and the resulting current for one phase is presented in Fig. 24.

In the switching pattern, it can be observed that there is a more low-frequency wave that is imposed on the machine, in this case having a frequency of slightly lower than 50 Hz. Due to that the machine windings contain inductance, it means that the high-frequency switching of the voltage will only lead to fairly small

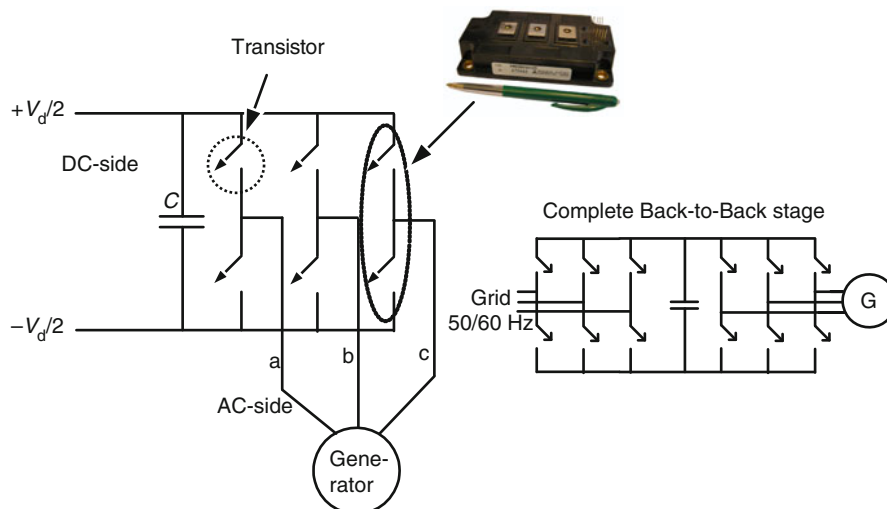
current ripple (added on the desired fundamental waveform) as can be observed in the figure. Since it is the current that forms the torque in an electrical machine, the torque produced by the generator will in principle only be based on this desired low-frequency signal. The higher switching frequency is filtered away.

The same power converter can also be connected towards the power grid, as shown in Fig. 23, however into the power grid it is not allowed to inject too much current harmonics. A problem is that the grid does not have enough inductance and accordingly a filter is needed in order to bring down the current harmonics to reasonable values.

The losses are fairly low in a power converter and can be down to 2% for a DC/AC-conversion in a wind turbine converter. However, there is also a need to convert the voltages and currents from DC to the grid AC-quantities, and accordingly around 3–4% is what can be achieved regarding converter efficiency. At low load, the efficiency is, of course, lower.

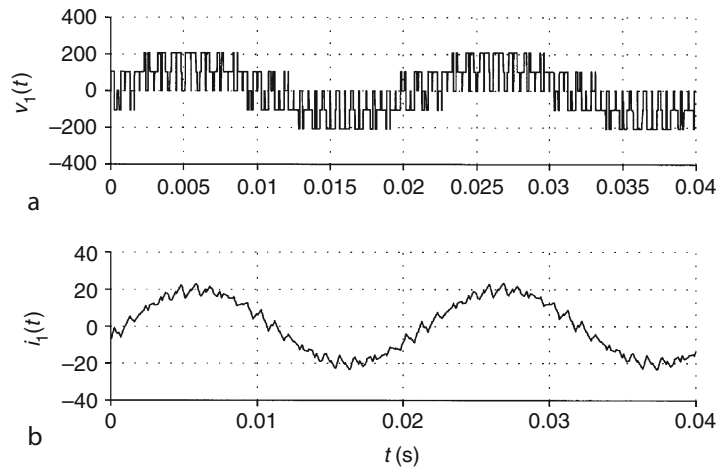
STATCOM and SVC

A STATCOM is in principle the same object as the converter presented in Fig. 24. However, on the DC-side there is only a capacitor. The AC-side is connected



Wind Power Generator Systems and Local Power System Interconnection. Figure 23

Left: Principle sketch of switching converter. Upper right: 200 A power module consisting of two transistors, making up the phase voltage *c*. Lower left: A complete B2B-stage, making the conversion from the generator voltage to the 50/60 Hz AC-voltage of the grid



Wind Power Generator Systems and Local Power System Interconnection. Figure 24

A phase voltage created by a three-phase converter as well as one resulting machine phase current

to the grid. It can be used to adjust the reactive power that is exchanged with the power grid. Typically it can be used to compensate a shortage of reactive power, either due to long inductive lines between the wind installation and a strong grid connection point, or due to that traditional fixed-speed wind turbines with insufficient reactive power compensation is used. Moreover, it can be used to stabilize the network in case of grid disturbances.

An SVC is a simpler and cheaper variant, which usually has a sufficient operational performance for performing the phase compensation.

Wind Turbine Generating Systems

A frequently asked question is: Which electrical generating system is the best, and this question is not possible to answer. The systems all have good performance and the differences in performance versus cost do not give a clear answer. However, it is obvious that there has been a shift from traditional fixed-speed turbines toward variable-speed turbines.

When it comes to produced energy per swept area given a certain average wind speed, all the concepts produce a similar amount of yearly energy [9], if really old-fashioned fixed-speed turbines, (which were not equipped with the two-speed facility as will be explained below) are disregarded.

Typical stator voltages are 690 V. However, the voltage levels are slowly moving upward as the power

increases further above 2–3 MW. A benefit of being below 1,000 V is that there are more available service technicians.

Fixed-Speed System

The fixed-speed wind turbine system was the wind turbine system that dominated the market until the millennium shift (2000). It is still used today in power levels up to 2 MW and is a well-proven design. Often this system is referred to as the “Danish concept.”

This type of wind turbine system is equipped with an induction generator which is directly connected to the grid. It is not a good solution to have a synchronous generator directly connected to the grid in this application since it will be too stiff, and cause large stresses on the drive train of the turbine (shafts, gearbox). The small slip that the induction generator has (0.5–1% at rated operation) will give sufficient smoothness.

Larger turbines of this type (>1 MW) are always equipped with the possibility to operate at two specific speeds. Either they are equipped with pole-changing generators which provides a lower generator output with a lower fixed speed and a larger generator rated power with a higher, fixed rotor speed. The turbine can also be equipped with two separate generators. In this way, the energy efficiency is as good for this turbine, given a certain rotor swept area and maximum shaft torque, as that of a variable-speed wind turbine. Of course, a variable-speed turbine is able to convert

more of the energy in the wind into shaft power, since it can exactly match the rotor speed to the wind conditions, [9], and in this way more of the power available in the wind can be converted into mechanical power. However, since no power electronics is used, there are no losses in a converter system. In addition, the two-speed facility still has a very high energy capture from the wind. Finally, the fact that the low speed generator has a lower flux level gives lower no-load losses, which is perfect for low wind speed operation. The turbines are as mentioned earlier, stall-controlled (the blades are mounted fixed into the hub, and cannot turn around their axes), more modern ones with active stall control (meaning that the blades turn slightly in order to be able to adjust the power level at high wind speeds, also adapting to temperature changes and different pressure conditions).

These types of turbines are robust and cheap, since the electrical system in principle only consists of an induction generator with shunt capacitors connected directly toward the grid as presented in Fig. 25. The pole-pairs of the generators are either 2 and 3 or 3 and 4 depending on the preferred speed range of the turbine. The first combination gives that the lower rotor speed is 67% of the higher one, and the second combination that the lower rotor speed is 75% of the higher one.

A slight drawback is that the starting of an induction machine requires large currents to magnetize the

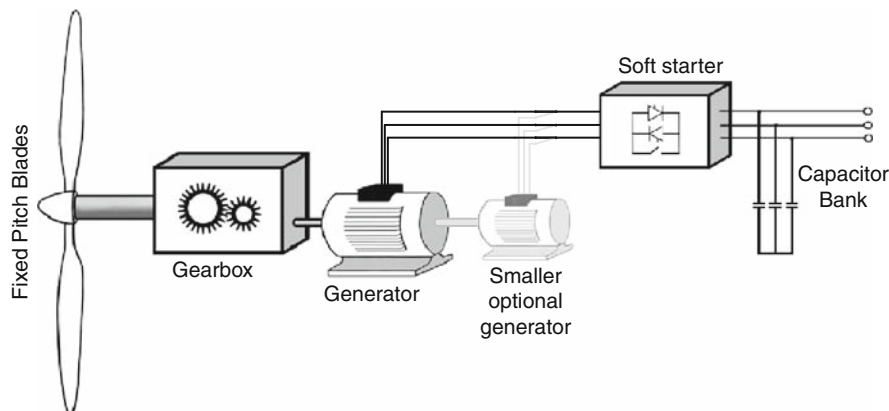
machine. In order to overcome this problem to some extent, the turbine is equipped with a soft-starter (consisting of antiparallel thyristors) in order to keep the currents below rated, when the turbine is being connected to the utility grid.

Phase-compensating capacitors are used to compensate for the no-load reactive power consumption of the generator, or in some cases also for full-load operation. In order to obtain a better controllability, an SVC or a STATCOM can be added in shunt with the wind energy unit.

Smaller units have the transformer connected outside the turbine while larger ones have the transformer either at the bottom of the tower or up in the nacelle. An advantage of having the transformer up in the nacelle is that it helps as counterweight towards the turbine rotor.

Fixed-Speed Systems with Variable Rotor Resistance

The fixed-speed turbine with a variable rotor resistance is in principle the same as the previous one, with reactive power compensation by capacitors and a soft-starter used to keep the currents below rated during the process of connecting the turbine to the grid, see Fig. 26. The difference is that the rotor resistance can be controlled which gives this turbine a limited rotor speed controllability and also fault current control



Wind Power Generator Systems and Local Power System Interconnection. Figure 25

Simplified electrical scheme of a fixed-speed wind turbine system. Here a two-generator system is presented, one larger with a higher rotational speed and higher power and one with lower speed and lower power. It is also possible to have so-called pole-shifting generators which can have these two generators in one generator

capability. The rotor speed can now be controlled in a speed range of about -2% to $+5\%$ from the nominal one. A difference toward the previously described fixed-speed wind turbines is that the fixed-speed wind turbine with a variable rotor resistance usually is pitch controlled.

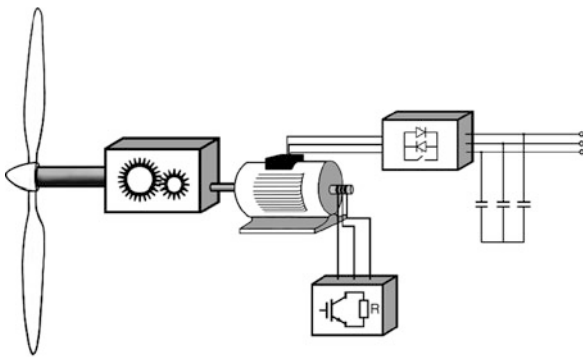
The fixed-speed turbine with a variable rotor resistance was manufactured by Vestas in the 1990s, but the production continued for many years for delivery to the US market.

Full-Power Converter Systems

The full-power converter system gives the highest degree of freedom regarding the control of the electrical

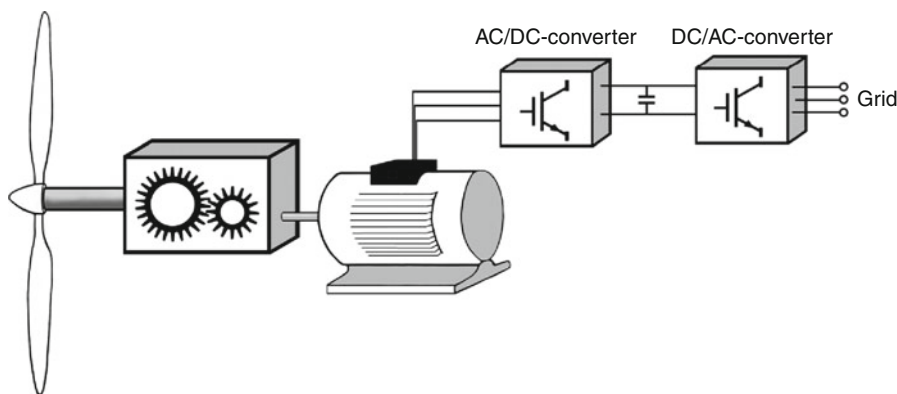
quantities as well as the mechanical ones. The name “full-power converter system” comes from the fact that all the produced wind power has to go through the converters, which is not the case for the DFIG system, which will be explained in a coming section. In this concept, the generator is completely decoupled from the grid by a power electronic AC/AC-converter between the stator of the generator and the grid. This means that from a grid point of view, it is irrelevant what type of generator is used, or if a gearbox is used or not. Siemens uses a cage-bar induction generator while General Electric uses a permanent-magnet generator. Both these systems are with an ordinary gearbox. Recently, more and more manufacturers are taking up this concept. Also, the full-power converter turbine is pitch controlled.

In this system, all the power is flowing through a DC-link in the converter. The converter consists of two sets of DC/AC-converters as the ones presented in Fig. 23. The reactive power is completely decoupled between the machine and grid side of the DC-link and can be produced or consumed separately on each of the DC-link sides. In fact, the grid side converter can be programmed to work as a STATCOM, in addition to its original task of feeding active power from the turbine into the grid, with very little additional cost. The amount of reactive power that can be supported to the grid is determined by the size of the grid side converter. This concept is presented in Fig. 27.



Wind Power Generator Systems and Local Power System Interconnection. Figure 26

Electrical scheme of fixed-speed system with a variable rotor resistance



Wind Power Generator Systems and Local Power System Interconnection. Figure 27

Simple scheme of full-power system

Slowly Rotating Generators Which Operate Without Gearbox

The objective of using a slowly rotating electrical generator, although this will be a much larger generator for the same power rating, is that the gearbox can be eliminated (or made smaller, since the gear ratio can go down), thus saving the cost of the gearbox and saving the gearbox losses, in addition to the increased reliability. As for many other aspects of wind energy, there is no clear answer if it is worth to use a slowly rotating generator and removing the gearbox or not.

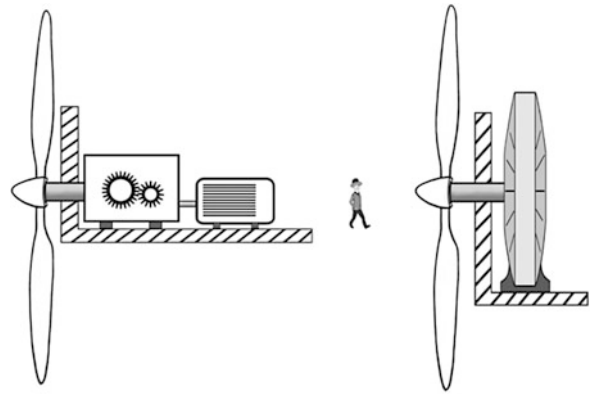
An AC-machine operates at a frequency governed by the frequency of the voltage applied to the generator (for the synchronous machine exactly the same, for the induction machine, almost the same, just the slip must be considered). However, as mentioned before, there is an additional design parameter that can give a certain ratio between the applied voltage frequency and the rotational speed of the machine and that is the pole-pair number. The relation between the applied electrical frequency and the mechanical speed as a function of pole-pair number is

$$\Omega_m = \frac{\omega_{el}}{p}$$

where Ω_{mech} is the mechanical rotational speed and ω_{el} is the electrical angular frequency. The drawback is that the large number of poles will make the machine to have a very high diameter. In order to reduce the needed number of pole-pairs and in this way keep the radius down, the supply frequency can be lowered to some extent in these systems. In Fig. 28, a sketch comparing the difference between the geared system and the non-geared system is presented. There is no difference in grid interface equipment (converters, etc.) between the two systems.

The directly driven machines can typically be supplied with a slightly lower electrical frequency, for instance 20 Hz, corresponding to 120 rad/s giving a needed number of pole-pairs of 100–200. For more information, the interested reader can in [10] find a comparison that is done based on a 500 kW system.

The largest manufacturer of this type of system is Enercon that uses electrically magnetized field windings. The consequence is that the rotor diameter and thus the whole generator diameter become very large. A possibility to have a rotor with a lower



Wind Power Generator Systems and Local Power System Interconnection. Figure 28

Principle sketch of conventional (geared) and directly driven generator, interface with converters to the grid has been omitted, but is the same for both systems

diameter is to use permanent magnets instead of the electrically magnetized rotor poles. In addition, the efficiency can be increased in this way.

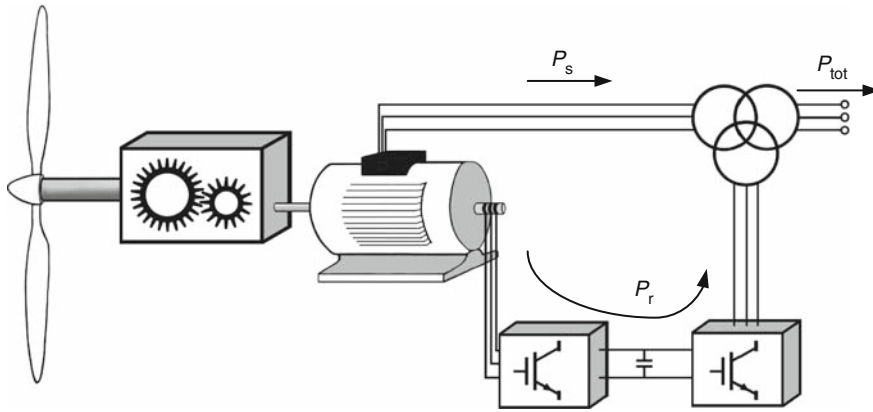
Doubly Fed Induction Generator Systems

During the last years (2005–2010), this has been the dominating wind turbine type. The converter in this turbine is located in the rotor circuit path giving a cheap and energy-efficient solution for variable-speed operation. The reason is that the converter rating only has to handle 20–30% of the turbine rating which both gives a lower investment cost and lower losses in the power electronic converter. A simplified electrical scheme of the system is presented in Fig. 29.

In the doubly fed induction generator, the currents in the rotor are formed by the rotor converter instead of only being induced from the stator. Of course, the induction from the stator must be considered when the rotor currents are being controlled. The rotor currents can now be controlled with a slip much larger than what is possible with a cage-bar induction machine. However, in order to handle the induction from the stator, the rotor voltage must exceed the slip (s) times the stator voltage

$$U_r > sU_s.$$

If this criteria is fulfilled, the rotor converter can feed in a current having a suitable frequency and in this



Wind Power Generator Systems and Local Power System Interconnection. Figure 29
Electrical scheme of DFIG system

way obtain a variable-speed operation. The power direction is actually from the grid into the rotor at low wind speeds, but at high wind speeds, the power flows from the rotor toward the grid.

It can actually be discussed if this is an induction machine. It is also possible to feed in DC-currents into the rotor, and then it is an electrically magnetized synchronous generator.

For a DFIG system, it is a trade-off between speed span and dimensioning of the rotor converter and rotor windings. The closer to the synchronous speed of the system that the turbine can be operated at, the lower the voltage applied to the rotor windings will be and also the lower the power transfer, in the rotor circuit including converter, in relation to the whole power can be.

Initially it can be thought that it would be favorable to have the synchronous speed of the system in the middle of the rotor speed interval. However, it is more favorable to have the synchronous speed a bit higher. The power that goes through the rotor in relation to the total power is proportional toward the slip in the machine.

$$P_r = sP_{\text{tot}}$$

It is accordingly important to keep the slip low at the highest power level. At lower power levels, it is OK with a higher slip since the transferred power anyway is fairly low. The fact that a variable speed wind turbine shall have a low rotational speed at low wind speeds in order to follow the optimal power coefficient, C_p , and

an upper speed limit gives a speed–torque operation region as presented in Fig. 30.

From the figure, it can be observed that a rating of 20% of the full power is sufficient for this turbine and the speed span is in principle 50–100%. A full-power converter system can operate a turbine at 0–100% of the speed; however, 40–100% is what typically is used, which means that the rotor speed range of the DFIG turbine is almost “full-speed range.” The investment cost and the converter losses are thus lower compared to the case if the whole energy flow would have had to be coped by a full-power converter. However, it should be mentioned that the rotor voltage is also proportional toward the slip and the stator voltage

$$U_R = s|U_S|$$

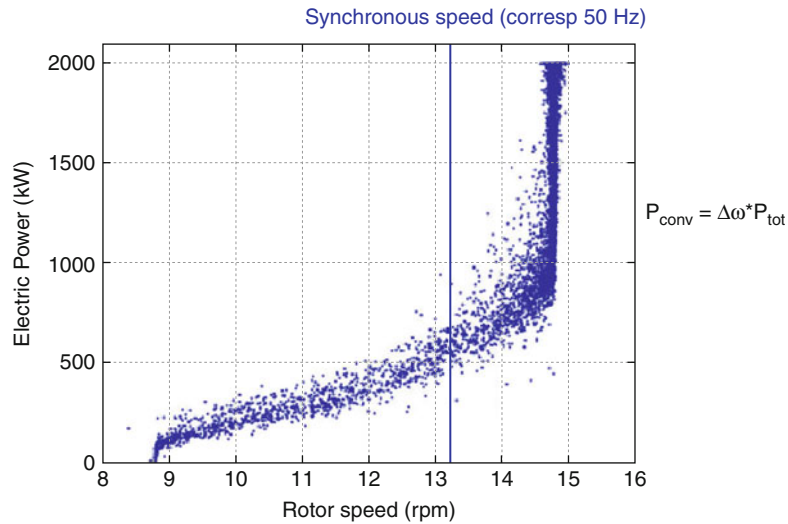
where

$$s = \frac{\omega_S - \omega}{\omega_S}$$

and where ω_S is the synchronous speed.

From the figure, with the location of the synchronous speed it can from the equations above be deducted that the highest rotor voltage will occur at the lowest rotor speed. Had the synchronous speed been in the middle of the speed span, the maximum rotor voltage could have been lower. However, a suitable trade-off is to select a bit higher synchronous speed as presented in Fig. 30.

Worth mentioning is about voltage levels of the rotor converter. It is beneficial to go up a bit in voltage,



Wind Power Generator Systems and Local Power System Interconnection. Figure 30

Rotor speed versus power relation for DFIG wind turbine. The rotational speed 13.2 rpm corresponds to the electrical frequency of 50 Hz, accounting for pole-pairs in the generator as well as the gearbox ratio. $P_{\text{conv}} = P_r$

for the used power level for the converter, since the currents then come down. The way that this is done is that the number of turns in the rotor typically is higher than in the stator. In this way, the voltage level of the rotor is still in the same vicinity as for the stator.

The energy efficiency of this system can be slightly improved by operating the generator in Y for low wind speeds and Delta for higher wind speeds [9]. For Δ -Y connections of an electric machine, please consult one of the electric machine textbooks.

The reactive power can easily be controlled to unity power factor using the DFIG system. The fact that the turbine is of variable-speed type, gives a smooth active power output. Since it is a variable-speed turbine, pitch control is used, as mentioned earlier.

A drawback with this system was earlier that it was difficult to have a good immunity toward severe voltage dips [11]. This is due to the fact that faults in the supply network creates large over-currents in the rotor for which the power electronic converter is not designed. Today (2011) solutions to handle these severe grid disturbances are provided by the manufacturers.

Power Quality Impact

The voltage in the grid should ideally consist of one sinusoidal component 50 or 60 Hz and nothing else.

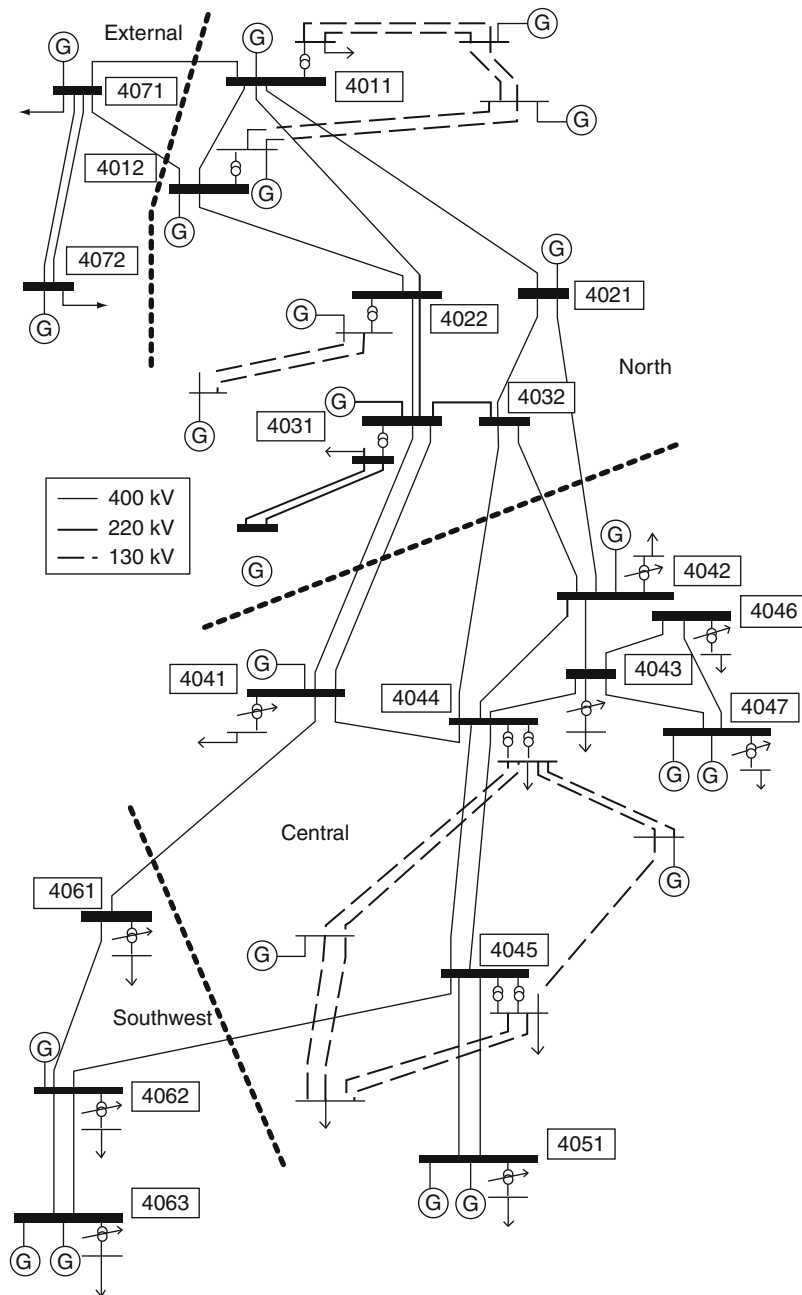
The magnitude of this voltage shall always be the same and in the three-phase system the voltages shall be separated 120° apart. In this case, the power quality is excellent. Any deviation, such as harmonics, voltage level variations, deviations in frequency, etc., is a negative power quality impact. In case the deviations become too large, then there is a risk for having power quality problem [7].

Traditionally, there has been a fear for negative power quality impact from wind turbines. However, this fear has more been based on uncertainties rather than on facts, in particular there has been a fear about flicker problems, but hardly any has been reported.

Anyway, in the beginning, the wind turbines did not have the best power quality impact, so they could be said to be passive grid disturbers. However, today modern turbines have a very good power quality impact, and they can even reinforce the grid. So, it can be said that wind energy converters have gone from being passive grid disturbers to being active grid-reinforcing units.

Power Quality Aspects

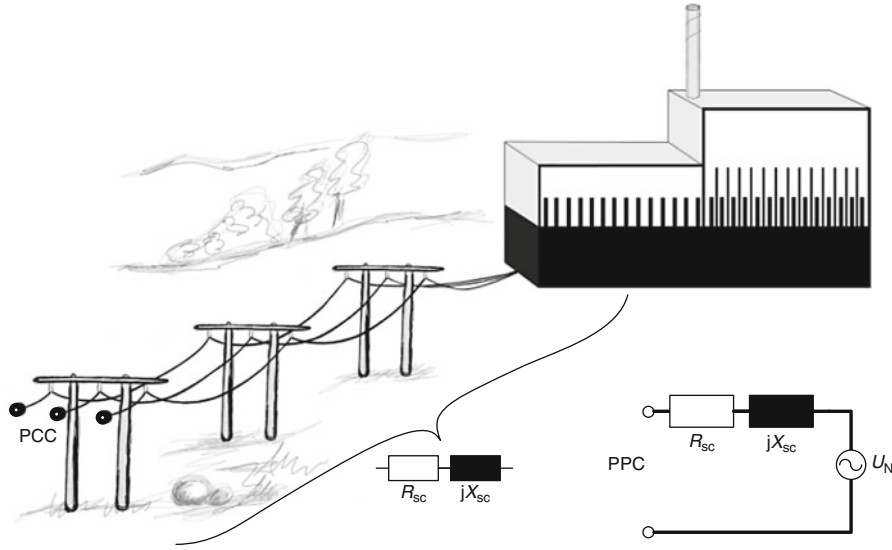
Definition of Equivalent Network A power system consists of a large amount of nodes connected together by power lines, as the one presented in Fig. 31.



Wind Power Generator Systems and Local Power System Interconnection. Figure 31
 Meshed network, in this case based on the Cigre Nordic 32-bus system

In order to perform calculations on the voltage level influence that a new power unit would have on the grid, as well as to determine possible fault currents, a network equivalent is constructed, consisting of an infinitely strong voltage source, and an impedance, according to Fig. 32.

An infinitely strong voltage source is used, in the picture it is here represented by a nuclear power station, but consists in principle of all objects connected to grid. From this infinitely strong voltage point, an impedance is defined to a grid connection point called PCC, Point of Common Connection. In case that there



Wind Power Generator Systems and Local Power System Interconnection. Figure 32

Definition of network impedance, that is, a line with an equivalent impedance and an infinitely strong voltage source behind the impedance, making up the PCC calculation point

would be a short-circuit in the PCC, a short-circuit current would appear with a value of

$$I_{sc} = \frac{U_N}{\sqrt{3}Z_{sc}}.$$

A commonly used term is the grid strength, and it is defined as

$$S_{sc} = \sqrt{3}U_N I_{sc}$$

often called short-circuit capacity of the grid, $\sqrt{3}$ is a factor originating from the calculations used for a three-phase system, but is not needed to comprehend in order to follow the calculation setup.

When a new power source, for instance a wind turbine installation is going to be connected to the grid, a grid strength and information of the grid impedance is obtained from the responsible utility, from which the equivalent resistance and reactance can be determined. The relations are

$$R_{sc} = Z_{sc} \cos(\psi_{sc})$$

$$X_{sc} = Z_{sc} \sin(\psi_{sc})$$

where ψ_{sc} is the grid short-circuit impedance angle, R_{sc} , X_{sc} , and Z_{sc} are the grid short-circuit resistance, reactance, and impedance, respectively.

Voltage Level Influence in PCC Now, if an energy production is connected to the PCC the voltage increase would be

$$\Delta U = R_{sc} I_{active} + X_{sc} I_{reactive}$$

where I_{active} and $I_{reactive}$ are the active and reactive powers coming from the energy production unit, for instance a wind farm.

The active and reactive current components can be found as

$$I_{active} = \frac{P}{U_N}$$

and

$$I_{reactive} = \frac{Q}{U_N}.$$

Worth noting is that the reactive power, Q , can be altered for variable-speed wind turbines and the voltage influence can then be controlled. Keeping Q to 0, gives lowest line losses and is what today (2011) typically is done, meaning the power factor, $PF = 1$.

The utility has to keep the voltage levels within certain limits. In Sweden, EN50160 states $\pm 10\%$ on lower voltage levels. Since other loads are connected to the lines as well as the fact that there is a need for

safety margins, a wind turbine installation could be allowed to vary the voltage level within a few percent. As can be noted from the equation above, the voltage level will increase as the power production increases. As mentioned above, the reactive power exchange with the grid can be used to control the voltage level. Today (2011) this option is rarely used, since the overall control of the power network could become more complicated if a lot of distributed generation tried to control the voltage levels in their vicinity.

Voltage Level Influence at the Turbine The wind energy installation could also have a certain distance to the PCC and then there will of course also be a voltage difference between the wind turbine installation and the PCC, see Fig. 33. The line resistance, reactance, and impedance are denoted as R_L , X_L , and Z_L .

The resulting voltage variation at the turbine compared to PCC then becomes

$$\Delta U = R_L I_{\text{active}} + X_L I_{\text{reactive}}.$$

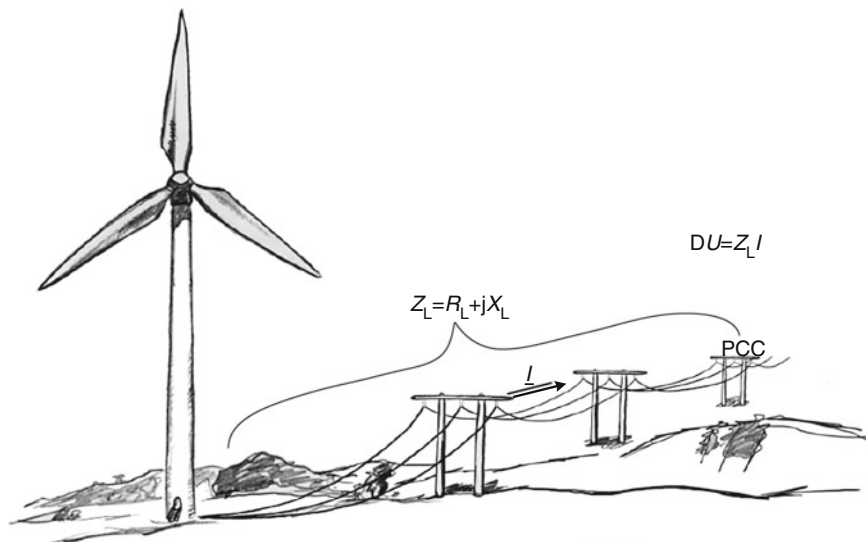
Voltage Fluctuations and Flicker Emission

The varying wind speed and also the mechanical design of a wind turbine cause a varying output power. This varying output power then causes voltage variations on

the grid to which the turbine is connected to. The observation from customers on a quickly varying voltage disturbance is that light bulbs start to blink. In [12], it is stated how the flicker evaluation should be conducted, based on experiments made more than half a century ago on when people started to detect a flickering light from an incandescent light. The result of using the flicker algorithm is the flicker severity index values, one 10 min value and one 2 h value, P_{st} and P_{lt} . If these values reach 1, most people start to get annoyed and some persons already can note a flicker level of 0.7. National regulations state how much flicker a wind turbine may contribute with to the grid. A typical value is 0.25 in P_{lt} . The sensitivity for a flickering lamp is as most sensitive for a frequency of 1000 changes per minute giving a frequency of 9 Hz, where only a 0.3% variation in the voltage magnitude can be observed. In Fig. 34, the sensitivity limit as a function of frequency is given.

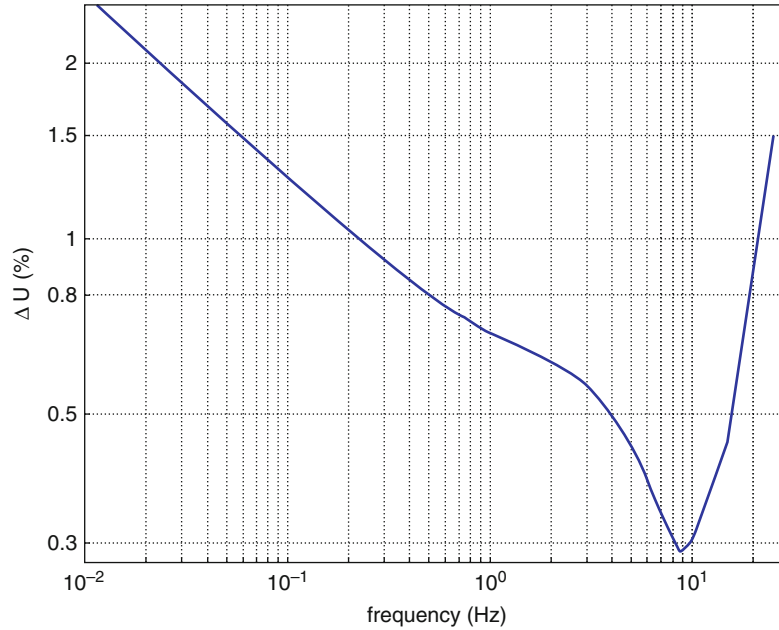
The standard IEC 61400–21 gives a guideline of how the flicker originating from a wind turbine shall be determined. The current coming from the wind turbine shall be recorded and fed into a fictitious network using various grid angles. The strength of the fictitious network ($S_{k, \text{fic}}$) shall be 50 times the rating of the wind turbine

$$S_{\text{sc, fic}} = 50 S_n$$



Wind Power Generator Systems and Local Power System Interconnection. Figure 33

Wind turbine attached to the grid as well as calculation of voltage level influence



Wind Power Generator Systems and Local Power System Interconnection. Figure 34
Limit for when a voltage variation can be detected by the eye

according to [13]. After having determined the fictitious flicker emission, $P_{st, fic}$ on this fictitious network, a dimensionless flicker coefficient c_f is formed as

$$c_f = 50P_{st, fic}$$

which then can be used to determine what flicker contribution a turbine will create on a grid to which it is being connected. The flicker coefficient shall be determined for four different grid short-circuit impedance angles, Ψ_{sc} , 30° , 50° , 70° , and 85° .

The development of modern variable-speed wind turbines has led to a situation where the flicker emission is not a problem today. In addition, when several turbines are connected in a group, these have a tendency to even out the power pulsations, and accordingly giving an even lower flicker emission. IEC 61400–21 states

$$P_{st} = \sqrt{N}P_{st, ind}$$

where $P_{st, ind}$ is the flicker emission from one single turbine and N is the number of turbines. However, according to [14, 15] the equation above underestimates the flicker emission from a whole wind park, in particular if traditional fixed-speed wind turbines are used. The reason is that the flicker emission does not

only originate from the wind turbines, also the grid itself imposes voltage disturbances on the turbines, which they reflect back as flicker emission. These reflected disturbances all come in the same phase from all wind turbines, and that is the reason why the equation above might give a bit too optimistic results.

A final comment is also that incandescent lights will slowly be outshifted due to their poor energy efficiency and in Europe this is driven by a legislative process that already is ongoing.

Harmonics Any period waveform can be composed by a number of sinusoidal components, according to

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos(n\omega t) + b_n \sin(n\omega t)$$

where

$$a_n = \frac{2}{T} \int_{t_0}^{t_0+T} f(t) \cos(n\omega t) dt$$

and

$$b_n = \frac{2}{T} \int_{t_0}^{t_0+T} f(t) \sin(n\omega t) dt.$$

In Fig. 35, an example of a square wave and its decomposition into Fourier components is shown.

Harmonics can occur both in voltage and in current. A wind turbine (or other components attached to the grid) can inject current harmonics into the grid. Depending on how strong the grid is, this will result in harmonics in the voltage. A strong grid will be less affected while a weak grid will be more substantially affected.

The voltage in the grid can never in reality be without harmonics, and these harmonics could disturb attached equipment, such as wind turbines. It is the responsibility of the grid owner to ensure that there are not too much harmonics in the grid.

The wind turbine in itself seldom gives rise to problems with harmonics. However, wind turbines connected to a local grid with cables, transformers, and inductors may as a combination amplify harmonics. Examples from literature [15] show that individual harmonics can be amplified at least three times. These local resonances is one concern regarding the failing transformers and generators in certain sea-based wind farms, so care has to be taken regarding PWM-patterns of switching converters and the operation of breakers, in particular vacuum breakers. For more details regarding PWM patterns, see Sect. Power Electronics – Converter. It should be pointed out that single wind turbines with PWM-converters provide sufficient filtering so without a local resonance there will not be a problem with harmonics.

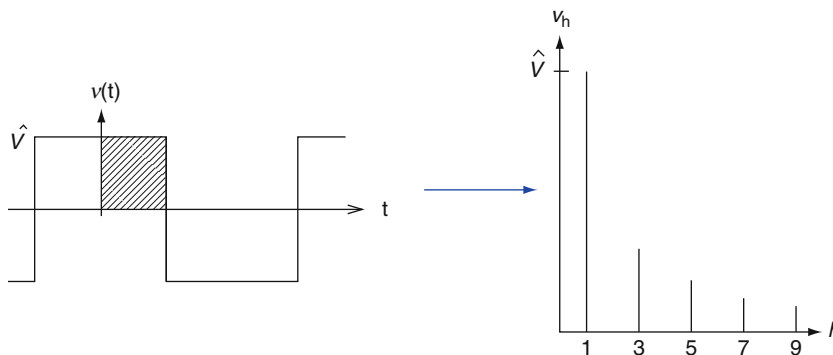
Fixed-Speed Wind Turbines

Steady-State Influence

Earlier, fixed-speed systems did not have a real reactive control ability. They had a capacitor connected to the stator terminal compensating for the no-load consumption of the turbine, which gave a voltage profile, depending on the network type, according to Fig. 36.

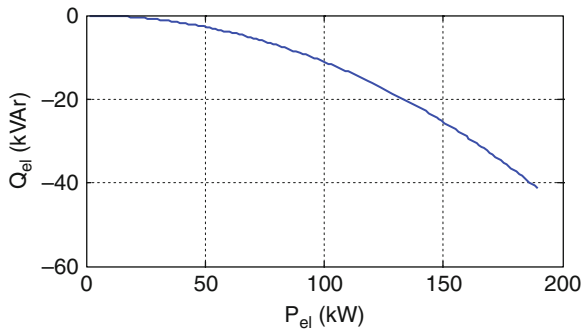
In a network with a grid strength (S_{sc}) of 20 times the strength of the connected network, the voltage level influence for three different network angles is presented in Fig. 37. In this figure, the reactive power at no-load that the induction generator needs, has been compensated for using capacitors. That is why the voltage change is zero for zero power. Otherwise, the produced power-voltage relation would be different.

Another point that needs to be addressed regarding Fig. 37 is that today capacitor banks are often connected in stages, which are used to keep the power factor close to unity for various operating conditions. This would mean that the voltages in Fig. 37 would increase somewhat. In principle, more capacitors could be connected and used for reactive power control and such solutions are offered by manufacturers of this type of turbine. Full-power converter systems and well as DFIG systems have an inherent possibility to control the reactive power and in the base case, the reactive power is kept as close to zero as possible. The result from a voltage level point of view is actually negative, since the reactive current no longer brings the voltage



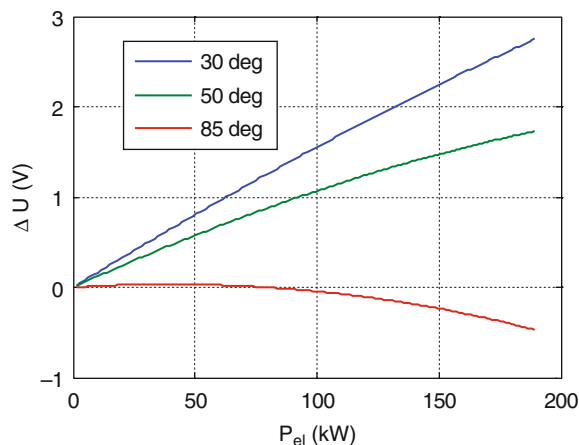
Wind Power Generator Systems and Local Power System Interconnection. Figure 35

Square wave and its decomposition into Fourier components



Wind Power Generator Systems and Local Power System Interconnection. Figure 36

Reactive power consumption as a function of active power for a traditional 180 kW wind turbine generator. Phase-compensating capacitors are installed compensating for the no-load consumption of reactive power by the wind turbine induction generator



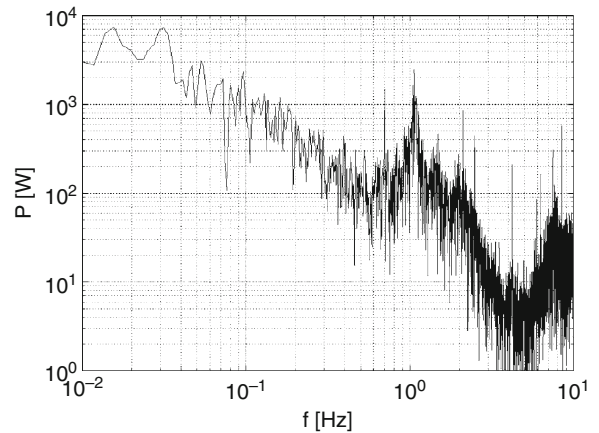
Wind Power Generator Systems and Local Power System Interconnection. Figure 37

Voltage level influence of a traditional fixed-speed wind turbine on a network with a grid strength of 20 times the rating of the wind turbines. Three different grid angles, Ψ_{sc}

level down, the voltage level will now be slightly higher given the same active power production.

Flicker Emission

Due to the inability of the fixed-speed systems to take up an incoming increase in the torque to the generator



Wind Power Generator Systems and Local Power System Interconnection. Figure 38

Output power spectrum of a fixed-speed wind turbine

by adjusting the rotor speed, these torque variations go through the system. In Fig. 38, the power spectrum from a fixed-speed turbine is presented.

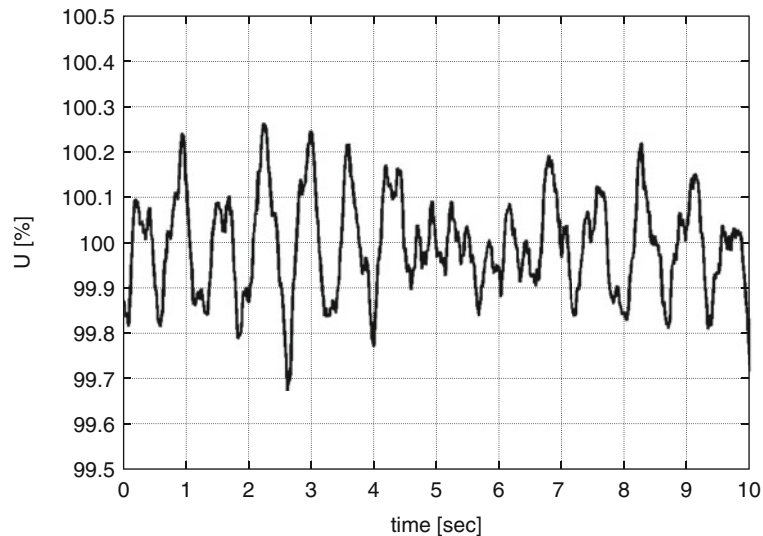
The reactive power will also vary and the consequence is of course that the voltage level on the adjacent network will vary. In Fig. 39, such a variation is presented close to a wind park consisting of three 180 kW turbines in a rural area.

In order to validate the severity of this pulsation, it is decomposed into its frequency components and presented as a function of fluctuation frequency together with the flicker limit line in Fig. 40.

There are several reasons for the unevenness of the power. The major reason is that the wind speed changes and thus the power production. In addition, the wind speed over the rotor disc is not constant, instead the wind speed across the rotor disc varies, and when the rotor rotates in this wind field, power pulsations are created. Figure 41 tries to illustrate this behavior.

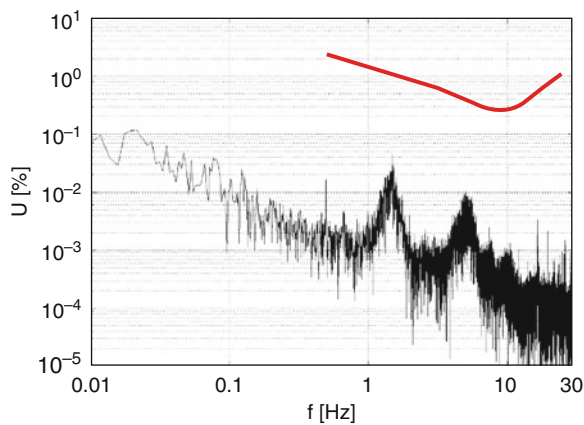
Apart from the variations that originate from the fluctuations that come from the wind, the turbine in itself as well as the generator also contribute to the output power spectrum.

Important to mention is that not only the turbine itself creates a flicker impact, the adjacent grid will also contain variations that through the turbine becomes flicker. Typically, this is the case for higher frequencies, [14].



Wind Power Generator Systems and Local Power System Interconnection. Figure 39

Voltage level variations close to an older fixed-speed wind turbine



Wind Power Generator Systems and Local Power System Interconnection. Figure 40

Voltage spectrum close to a fixed-speed turbine and in red the flicker limit

The flicker coefficients were earlier up to about 8 for fixed-speed turbines, but today the flicker coefficients are usually lower.

For a variable-speed turbine, the incoming mechanical power has the same behavior. However, a very important aspect is that for the variable-speed turbines, this torque/power variation is taken up

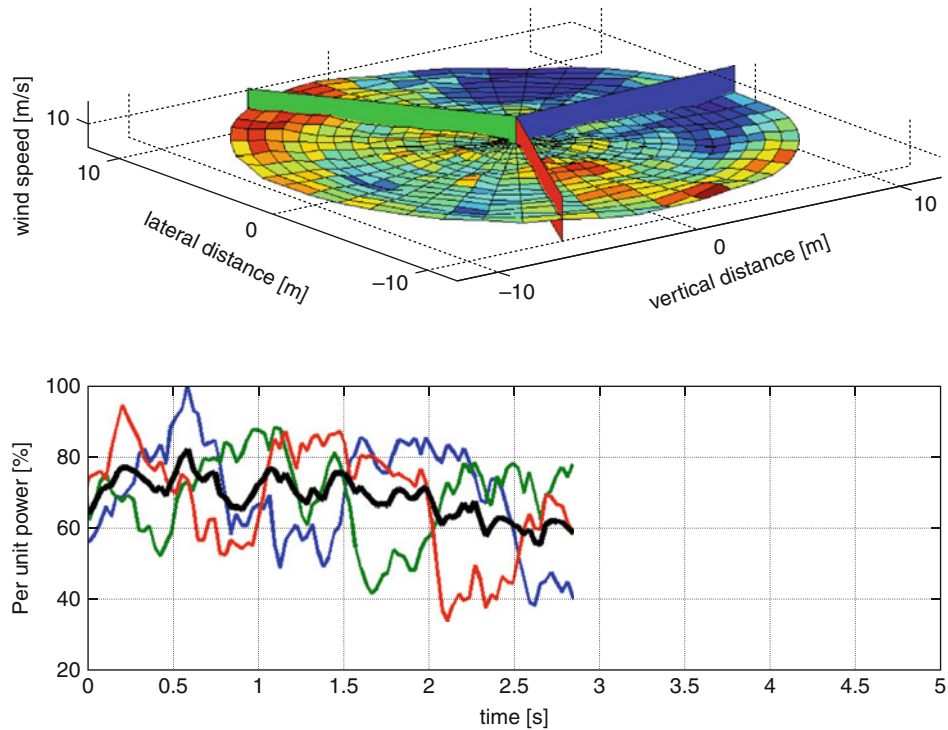
by small rotor speed adjustments, and the power pulsations do not come out to the grid. Accordingly, these “rotational-formed oscillations” are not visible in the output power from variable-speed turbines.

Special Considerations for the Start-up of a Fixed-Speed System

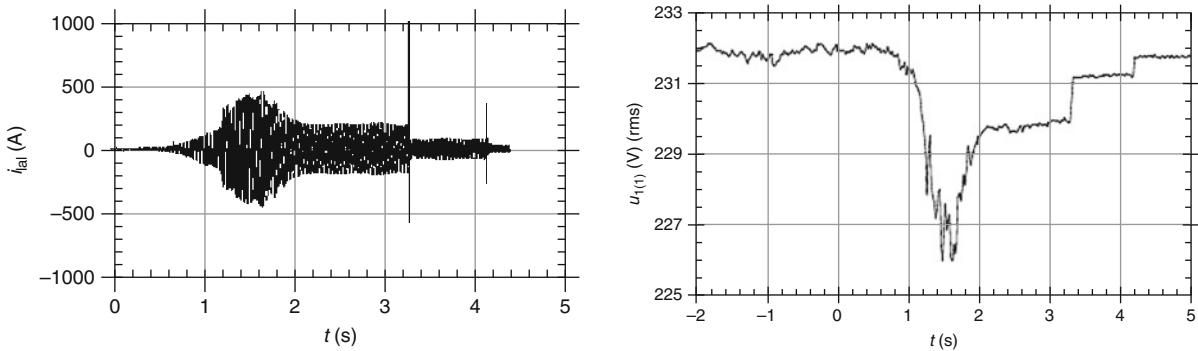
When a fixed-speed wind turbine is to be connected to the grid, the generator needs to be magnetized, which gives rise to a high current, seen in Fig. 42 below between 1 and 2 s.

Thanks to the soft-starter, the current is limited to a value below the rated current which in this case was a peak value of 400 A. Between 2 and 3 s the no-load-current of the generator can be observed. Then at the time instants 3.3 and 4.2 s two capacitor stages are connected and the current finally almost become zero. The consequence on the voltage level during the connection then becomes as presented in Fig. 43.

In IEC-61400-21 it is stated that a voltage change factor and a flicker step factor shall be determined. For variable-speed turbines, these coefficients become rather unimportant; however, for fixed-speed turbines they can be a limiting factor regarding the installation of such a turbine on a weak network.



Wind Power Generator Systems and Local Power System Interconnection. **Figure 41**
Blade rotation and resulting torque forming on the individual blades as well as the average torque formation



Wind Power Generator Systems and Local Power System Interconnection. **Figure 42**
Inrush current when connecting a traditional fixed-speed wind turbine

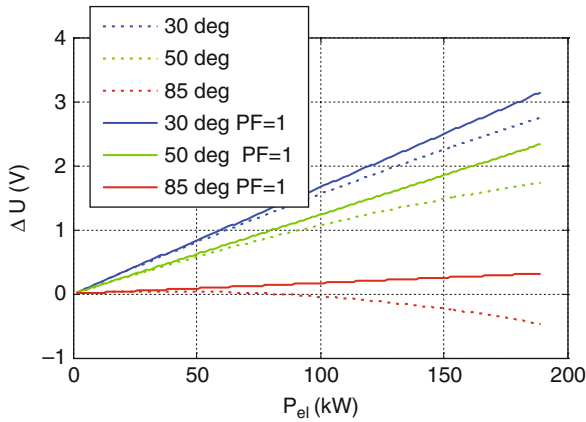
Wind Power Generator Systems and Local Power System Interconnection. **Figure 43**
Voltage level during the connection of a traditional fixed-speed turbine, ratio between P_n/S_{sc} is around 40

DFIG and Full Converter Turbines

Steady-State Influence

The variable-speed systems have an excellent reactive power control ability which can be used for voltage

control, but it is seldom utilized today. Using a power factor (PF) of one, the voltage level influence then becomes as presented in Fig. 44, using a grid strength 20 times the rating of the wind turbine. However, by using a variable power factor, as the traditional



Wind Power Generator Systems and Local Power System Interconnection. Figure 44

Voltage level influence by a fixed-speed (*dashed*) and variable-speed (or modern fixed-speed) wind turbine (*solid*). The fixed-speed turbine is assumed to have a variable power factor while the variable-speed turbine is assumed to have a constant power factor of 1

fixed-speed turbines, at least the older ones do, the voltage variations are actually lower, as seen in the figure. As long as the maximum current level of the turbine or line is reached, it is possible to completely eliminate the voltage variations due to the varying power production, by an appropriately selected reactive power injection.

If a system is designed to have an extra reactive power ability, apart from the voltage level control, this can also be useful during fault situations, and the electrical system must in some parts be rated to handle a higher current.

In the DFIG system, reactive current and thus reactive power can easily be generated through the generator, while the full-power system needs a larger grid side converter to produce reactive power for the grid. This is especially true when the wind turbine is operating at full power, then there is in the original design no room for reactive power production. For a full-power converter system an over dimensioning of the grid side converter by 8% gives a possibility to produce or consume more than 40% reactive power also at rated power operation. One large manufacturer of full-power converter wind turbines has a grid side converter that has a 40% over current capacity. For the DFIG it is also possible to generate reactive power, however, not

as easily using the grid side converter as for the full-power converter system.

At lower wind speeds, the turbine can consume or produce reactive power without an over-dimensioned grid side converter.

An interesting possibility is thus that a DFIG or full-power converter wind turbine system can act as a STATCOM when the wind turbine is standing still due to low wind.

Flicker Emission

Variable-speed turbines have the ability to control the rotor speed. In this way, the disturbances that originate from the wind as well as from the mechanical design of the turbine can be taken up as speed variations. In case the speed variations become too large (in particular that the rotor speed becomes too high at high wind speeds), the pitch control can be used to take the turbine back to proper rotor speeds. In addition, the converter also makes it possible to reduce the wind generating system's response to disturbances that are already existing in the grid. The result is that variable-speed turbines have flicker coefficients, c_f , of between 1.5 and 4, define depending on the grid impedance angle and average wind speed.

Active Power Control for Wind Turbines

Normally wind turbines are not used for load leveling, since this would mean loss of produced energy. However, with the exception of fixed-speed turbines without pitch control mechanism, wind turbines can be used for this purpose (as spinning reserve), but, this requires that the turbines are not operating at their maximum power production point, but instead with a reduced value. In the case that the network needs more power, the production can now be increased to the maximum one.

Already, grid codes, for instance that of ELTRA, suggest that this feature should be included in the control of a wind turbine, at least if it is a part of a large wind farm.

Harmonic Current Emission from Wind Turbines

Fixed-speed turbines emit harmonic currents mainly because voltage harmonics exist in the grid and these

are “mirrored” in the current coming out of the generating system. In addition, the electric generator in itself also contributes with harmonic currents, since the design cannot be completely sinusoidal. However, these harmonic current emission levels are usually very low.

Converter-operated wind turbines, of course emit harmonics, since they are equipped with power electronic converters.

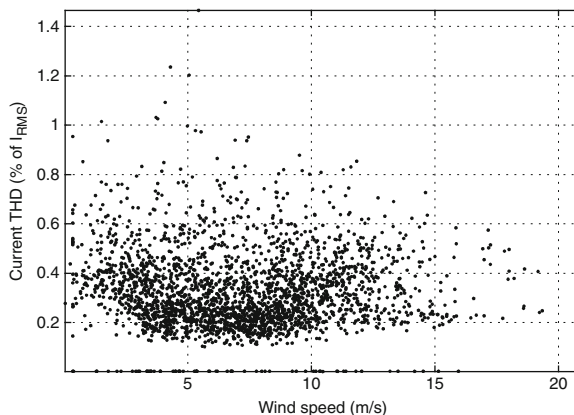
Full-power converter systems typically inject harmonic currents having a frequency of one or a couple of kHz. The turbines are equipped with harmonic filters, so that the current emissions are below the ones stipulated in the standards. Figure 45 depicts the harmonic current emission from a wind farm consisting of seven variable-speed wind turbines.

From the figure, it can be noted that generally the total harmonic current distortion is below 1.3% for this wind farm, which is a level far below usual limits.

Behavior During Network Faults

Fault Behavior and Requirement on Fault Behavior

Also during deviations in voltage and frequency, the wind turbines should continue to stay on line. For various regions, the valid grid code put requirements of the immunity level toward voltage and frequency deviation that the wind energy installation should be able to cope with. One example, however valid for



Wind Power Generator Systems and Local Power System Interconnection. Figure 45

Total harmonic current distortion (THD_i) from a wind farm consisting of seven variable-speed wind turbines

a whole wind farm connected to a high voltage level from E.ON Netz is shown in Fig. 46.

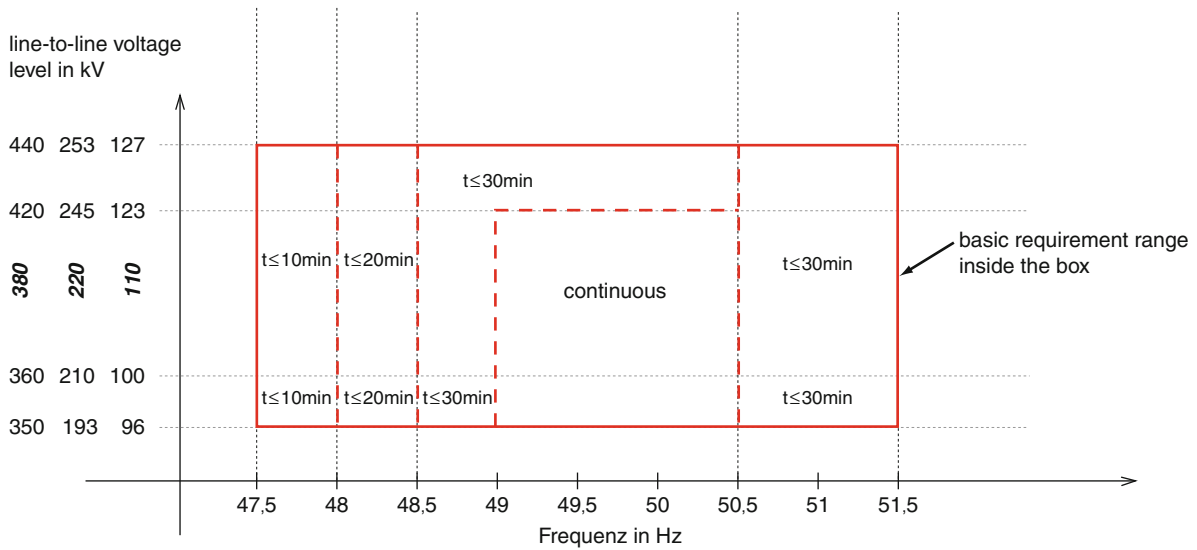
Although this requirement is for a large wind farm, at the end it still puts the demands on the individual wind turbines.

For each country, there is a specific grid code that states a reference voltage dip curve which future wind energy installations must be able to handle (Ride Through). Some grid codes specify the voltage curve at the PCC (Point of Common Connection, which means where the wind turbine installation is connected to the grid) while others, for example, the Swedish, states certain points in the main national network, in the Swedish system, the meshed 400 kV-system. If the grid codes specify the voltage behavior immunity level at PCC, it means that potential wind energy installation installers accordingly must do more advanced calculations (dynamic simulations) to check that their installation will meet this requirement. Depending on the various regions, there are different demands that shall be fulfilled. In Fig. 47, some examples are given.

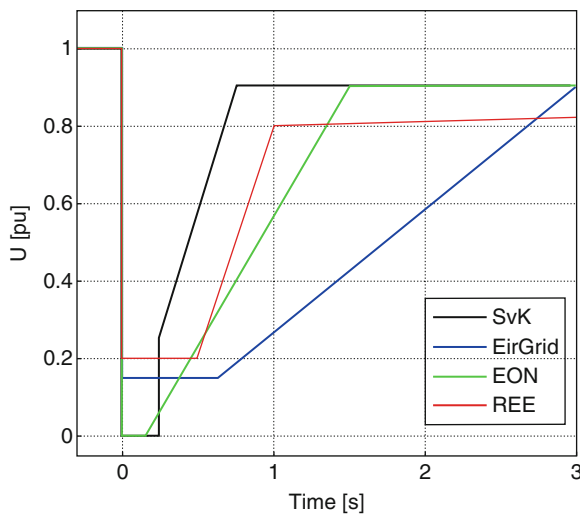
Fixed-Speed System

The response of a fixed-speed wind turbine is much dominated by the direct connection between drive train, generator, and grid. As induction generators in general, a reduced grid voltage may lead to excessive reactive power consumption by the turbine if the wind speed is somewhat high. In addition, there is a risk that a fixed-speed turbine can overspeed in case of a too low voltage, due to the fact that the electromagnetic ability of an induction generator deteriorates quickly at lower grid voltages.

Whenever there is a quicker change in the supplying voltage, the generating system of the wind turbine reacts strongly. The fixed-speed system contributes with a fault current according to an induction motor, but driven with a negative load. Again, the system has a limited ability to control the fault currents. In Fig. 48, the measured response to a fairly small voltage disturbance for a fixed-speed wind turbine is presented. As can be noted in the figure, the resulting electric output power oscillations range up to rated power, although the voltage dip is not unusually large, only about 10%. The real consequence is if the voltage adjacent to the



Wind Power Generator Systems and Local Power System Interconnection. Figure 46
Operational requirements regarding staying on line as a function of voltage level and frequency



Wind Power Generator Systems and Local Power System Interconnection. Figure 47
Examples of required under-voltage immunity for wind energy installations

turbine is influenced in such a way that it would disturb customers. And, in the example presented in Fig. 48, the turbine response would NOT cause any noticeable problems for nearby customers.

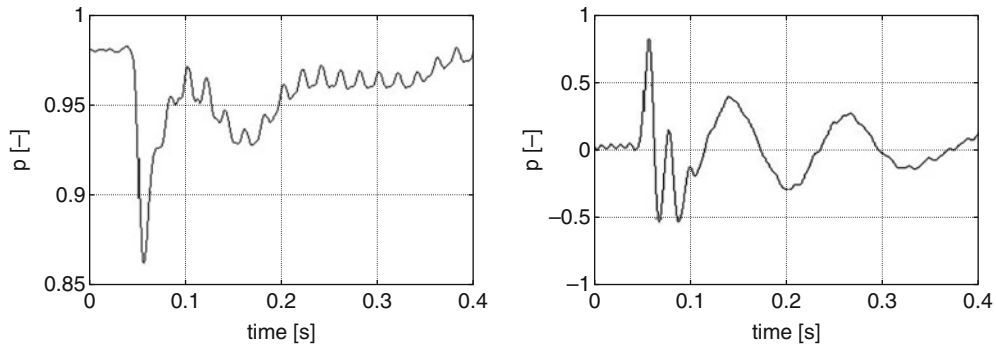
No recording of a short-circuit for fixed-speed wind turbine was available. In Fig. 49, the result from a 15 kW

machine exposed to a three-phase short-circuit is presented. In the figure, the currents are presented instead of the power. The peak currents are about seven times the rated ones. In this case, the voltage on the adjacent grid is anyway very severely disturbed, so the currents taken by the induction machine do not deteriorate the voltage quality for the nearby customers further. However, the wind turbine would be severely disturbed and would have to be disconnected from the electrical network. Overspeeding and overcurrent could damage the fixed-speed turbine. However, newer regulations today about under-voltage immunity, put harder restrictions on the wind turbines, also fix-speed ones, regarding whether the turbine is allowed to disconnect or not.

Earlier fixed-speed systems could not ride-through voltage disturbances. However, today a large manufacturer of a fixed-speed system provides a ride-through facility that is sufficient for the whole of Europe.

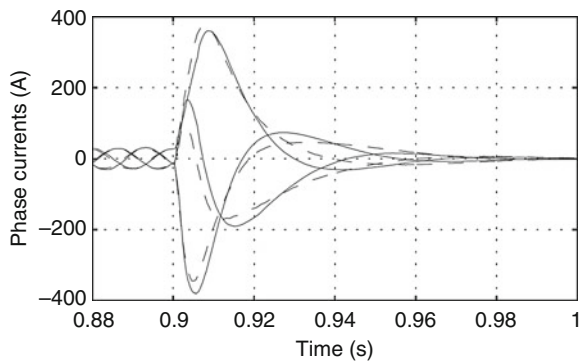
Fixed-Speed System with Active Pitch Mechanism

The fixed-speed systems with active pitching mechanism can to some extent use the pitch controller to provide a varying output that can be used to counteract a need from the grid. It is possible to change from 100% to 20% production within a couple of seconds. Till date, the turbines have not been designed for power/frequency control.



Wind Power Generator Systems and Local Power System Interconnection. Figure 48

Voltage magnitude and resulting electric power response for a fixed-speed turbine



Wind Power Generator Systems and Local Power System Interconnection. Figure 49

Currents from a 15 kW machine following on a three-phase short-circuit, solid lines are measurements and dashed lines are simulated values

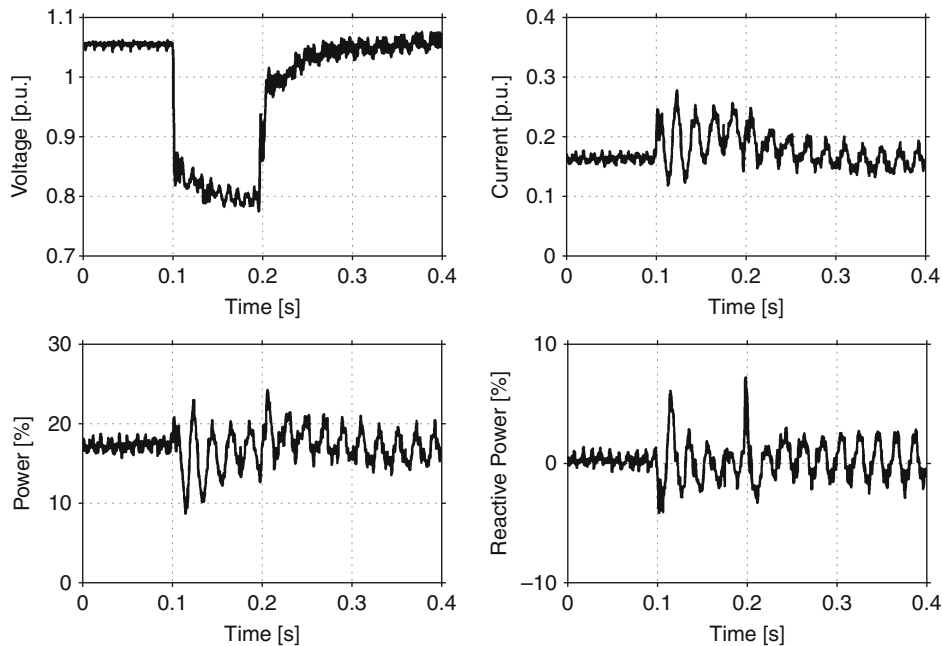
DFIG System

The immunity toward grid disturbance is for the standard DFIG turbine low, due to the small rotor converter used in this system. At a larger voltage dip, the rotor currents become too large. Only dips with a remaining voltage higher than 75–85% could be handled. In the original DFIG system the fault currents were similar or worse than for a fixed-speed system. When a larger fault occurred, this turbine had to quickly (within about 40 ms) be disconnected from the network. In Fig. 50, the response to a smaller voltage dip is presented and in Fig. 51 the response to a short-circuit. The smaller voltage dip is as noted, very well handled by the DFIG system, while the short-circuit, the turbine cannot handle.

In order to improve the immunity level to deal with the new grid codes, additional equipment must be added.

When the turbine is not allowed to be disconnected due to a disturbance, there is a surplus of power that will lead to overvoltages in the power electronics. The surplus of power is caused by the fact that when the voltage is low, the power that can be fed out into the electrical network is limited, since the current cannot be higher than the rated one. If the voltage is 0, no power can be fed out to the network. Then two things can be done. One is that the turbine is allowed to speed up a bit while the pitch goes into action and reduces the incoming power. When the voltage returns, the turbine can be pitched back to its pre-fault operating condition. However, this takes a while which is undesirable since from a power network stability point of view, it is desirable to have the pre-fault power production immediately when the voltage returns. Instead an alternative approach can be used. If the incoming power is kept constant, then the difference toward the power that can be fed out into the network can instead be dissipated by the turbine. For this purpose, typically power-electronic controlled resistors “breaking choppers” can be used.

Changes that can achieve this functionality are the insertion of power-electronic controlled resistors in the power electronic circuits, as mentioned above, or by using over-dimensioned power semiconductors. In Fig. 52 below a crude scheme for a ride-through system for a DFIG turbine is presented. A key component in the ride-through systems is the “breaking chopper” where the power that cannot be supplied into the grid



Wind Power Generator Systems and Local Power System Interconnection. Figure 50

Measured response to a minor voltage dip by a DFIG system

during the disturbance is dissipated. The breaking chopper can also be located on the AC-side.

In Fig. 53, a photo of the DC-link chopper is presented for such a system.

A photo of the AC-side crowbar for a wind turbine equipped with under-voltage immunity functionality is presented in Fig. 54.

These additions can nowadays (2011) be considered as state of the art. In Fig. 55, a recorded voltage waveform on a 2 MW DFIG system with a ride-through ability function is presented.

The current through the breaking chopper has the appearance as presented in Fig. 56 during the fault.

The resulting active and reactive power response is depicted in Fig. 57.

From the figure, it can be noted that the turbine very well manages to handle this voltage disturbance.

Full-Power Converter System

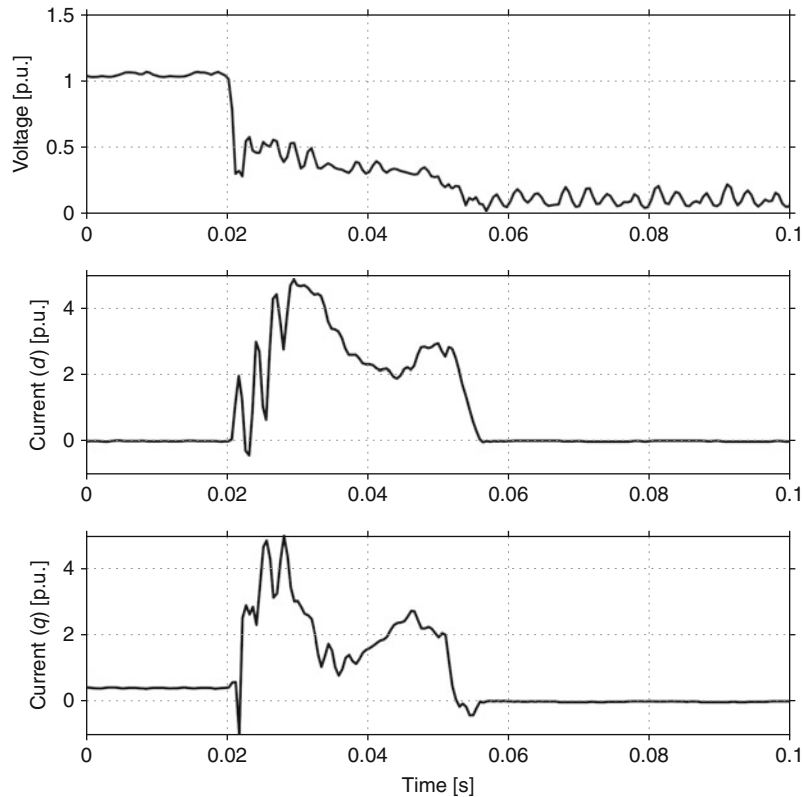
Any suitable fault current can be provided as long as the current rating of the converter valves is not exceeded. Observe that the contribution to the short-circuit

power of the grid usually is much lower for a variable-speed wind turbine than for a conventional synchronous generator connected directly toward the grid. This means that since a modern wind turbine with full-power converter has a very low fault current, it will be required to perform a review of the protection settings if such turbines are to be installed into a system in great numbers.

Comparison of Systems

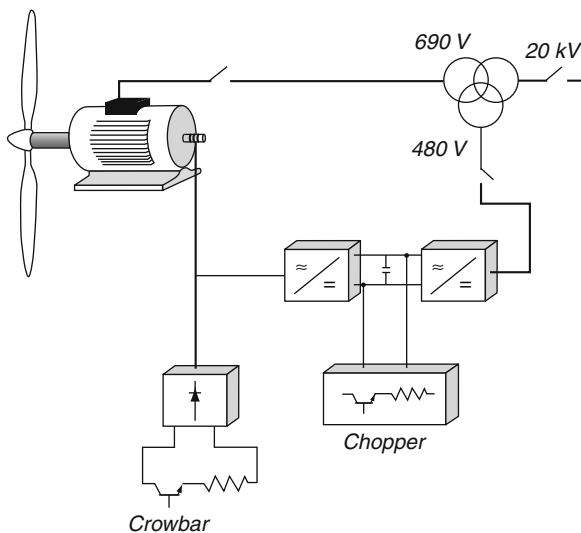
Comparisons of the various systems can be found in literature, so they will not be repeated here, and also the fact that in order to improve the grid performance several improvements have been made during the years, blending the properties of the system, leads to the fact that no such comparison is shown here.

The most important question is “What is the cost for each produced kWh for a unit that fulfills the grid codes,” and all systems can be made to fulfill the requirements, so it is mainly an issue of cost. For all systems there exist technical solutions, often using power electronics to obtain the desired performance.



Wind Power Generator Systems and Local Power System Interconnection. Figure 51

Measured response to a larger voltage dip (a short-circuit in the feeder) by a DFIG system



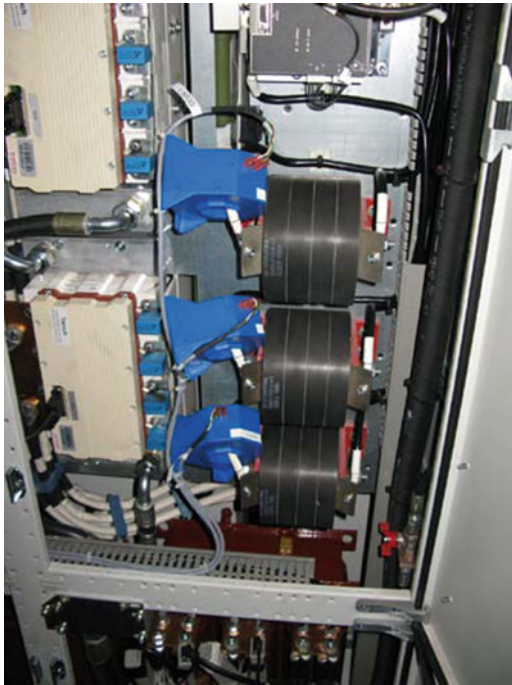
Wind Power Generator Systems and Local Power System Interconnection. Figure 52

Example of a ride-through system of a DFIG wind turbine system

Future Directions

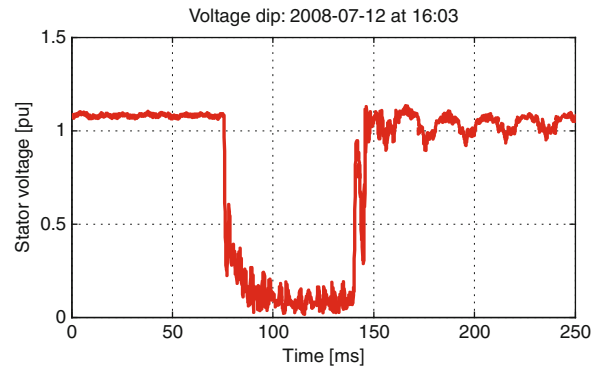
It is never easy to predict what will happen in the future, but some qualified guesses are possible to make. For sure the wind turbines will continue to increase in sizes, but where the upper economical size will be in the future, is not easy to predict. Moreover, the trend that more and more wind turbines will be power electronic driven seems also to be continuing. The reduction of losses in power electronic components could lead to the condition that the full-power converter system could be more favorable compared to the DFIG system. The permanent-magnet multipole generators seem from a research point of view to have a great potential; however, they today (2011) seem not yet to have reached industrial readiness at full scale.

Today (2011) there exist usually generating systems with a full-range gearbox (around three gearing stages) or no gearbox with a machine with a high number of poles. A possible development in the future is



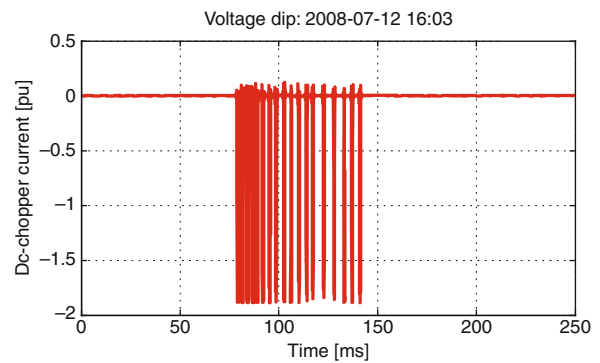
Wind Power Generator Systems and Local Power System Interconnection. Figure 53

Breaking resistors on the DC-link, black round “cans” in the left part of the picture. Blue objects are current sensors and the white plastic covers contain the transistors used to control the current through the breaking resistors



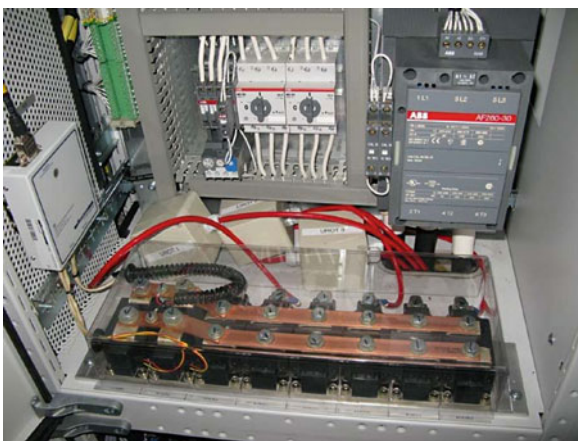
Wind Power Generator Systems and Local Power System Interconnection. Figure 55

Voltage waveform on July 3, 2008, on a DFIG turbine



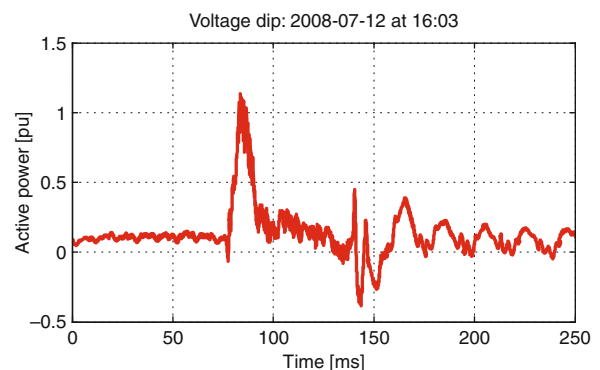
Wind Power Generator Systems and Local Power System Interconnection. Figure 56

Current through the breaking chopper



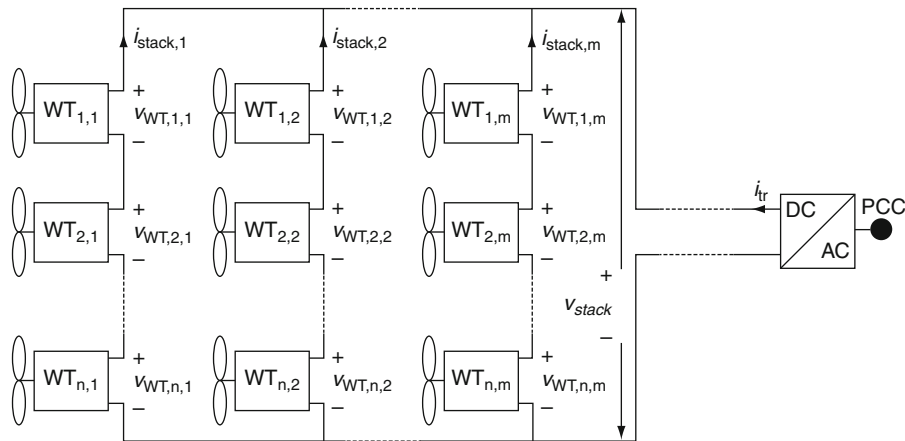
Wind Power Generator Systems and Local Power System Interconnection. Figure 54

AC-crowbar, at the bottom of the picture

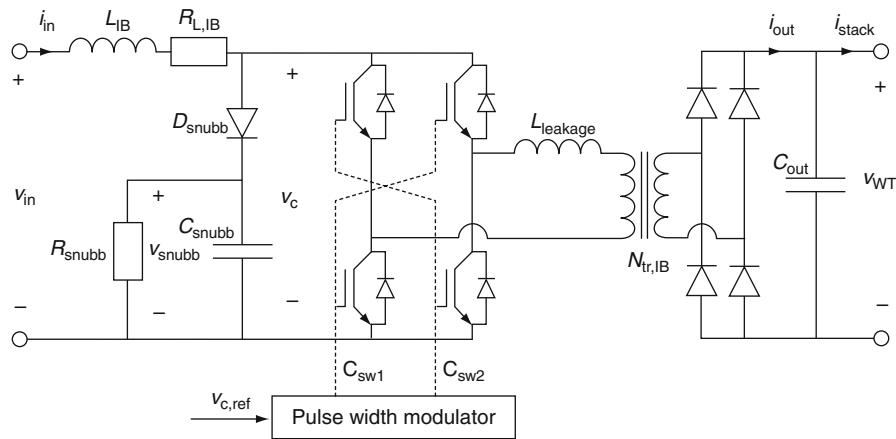


Wind Power Generator Systems and Local Power System Interconnection. Figure 57

Active and reactive power response due to a voltage dip down to almost 0% remaining voltage



Wind Power Generator Systems and Local Power System Interconnection. Figure 58
Series DC wind park



Wind Power Generator Systems and Local Power System Interconnection. Figure 59
Full bridge isolated boost converter suitable for a wind turbine with DC as output

that machines and gearboxes could become more integrated, and also perhaps one or two gearing stages could be combined with machines having an intermediate number of poles.

A possible future that means greater changes compared to the ones that have been mentioned above, is that HVDC-networks might be a reality in future sea-based wind farms. In such a case, the generating systems should be designed in such a way that they feed out DC in an appropriate way. In [16] a system based on wind turbines connected in series on their DC-outputs is presented according to Fig. 58.

In Fig. 59, an example of how the converter system for such a “DC wind turbine” could look like is presented.

Abbreviations

AC	Alternating current
DC	Direct current
IM	Induction machine
PE	Power electronics
PF	Power factor
THD	Total harmonic distortion

THD_i Total harmonic distortion for the current
X(n) Harmonic number

List of Commonly Used Variables

A Area [m²] real quantity
B Magnetic flux [T] imaginary quantity
C Capacitance [F]
F Force [Nm]
C_p Aerodynamic power coefficient
I Current [A]
I Complex current vector [A]
I_{active} I_{reactive} Active (quadrature axis) and reactive
 = **I_q I_d** (direct axis) current [A]
J Inertia [kgm²]
L Inductance [H]
N Number of turbines
N₁, N₂ Number of turns in a transformer
P_{conv} Power transmitted by a converter [W]
P_e P Electric power [W]
P_{st} Short-term flicker severity index
P_{st,ind} Short-term flicker severity index
 individual turbine
P_{lt} Long-term flicker severity index
P_{loss} Power loss [W]
P_{mech} Mechanical power [W]
P_R, P_S, P_{total} Rotor, stator, and total power [W]
P_w Available power in the wind [W]
Q_e Q Reactive current [VA]
R Resistance [Ω] radius [m]
R_L Line resistance [Ω]
R_{sc} Short-circuit resistance [Ω]
S Slip
S Apparent power [kVA]
S_{sc} Short-circuit capacity [VA]
S_{sc fic} Fictitious short-circuit capacity [VA]
S_n Rating of wind energy installation
 [VA]
T Torque [Nm]
T_e Electrodynamical torque [Nm]
T_s Shaft torque [Nm]
T_w Wind torque [Nm]
U Voltage [V]
U Complex voltage vector [V]
U₁, U₂ Primary and secondary voltage [V]
U_N Rated voltage [V]
U_R, U_S Rotor and stator voltage [V]

$\hat{U}$ U_{rms} Peak and rms-value of voltage [V]
V_{dc} DC-link voltage [V]
W Energy [Ws]
WS Wind speed [m/s]
X Reactance [Ω]
X_L Line reactance [Ω]
X_{sc} Short-circuit reactance [Ω]
Z_L Line impedance [Ω]
Z_{sc} Short-circuit impedance [Ω]
c_f Flicker coefficient
e Induced voltage [V]
i_r i_s i_t Phase currents [A]
n Rotational speed [rpm]
p Instantaneous power [W] pole-pair
 number
r Distance [m]
t time Time [s]
u_r u_s u_t Phase voltages [V]
ΔU Voltage variation [V]
Φ Phase angle between voltage and
 current [Degree]
Ψ Flux [Wb]
Ψ_{sc} Short-circuit angle [Degree]
Ψ_s, Ψ_r, Ψ_m Stator rotor, magnetizing flux linkage
 (“flux”) [Wb]
Ω_m Rotor speed [rad/s]
δ Load angle [deg]
λ Tip-speed ratio
ω Speed [rad/s]
ω_{el} Electrical speed [rad/s]
ω_s Synchronous speed [rad/s]

Bibliography

Primary Literature

1. Vestergaard J, Brnadrup L, Goddard RD (2004) A brief history of the wind turbine industries in Denmark and the United States. In: Academy of International Business (Southeast USA Chapter) conference proceedings, November 2004, pp 322–327
2. Dorf RC, Svoboda JA (2004) Introduction to electric circuits, 6th edn. Wiley, New York. ISBN 0-471-44795-1
3. Mohan N, Undeland T, Robbins W (2003) Power electronics. Wiley, New York. ISBN 0-471-22693-9
4. Cheng DK (1993) Fundamentals of engineering electromagnetics. Prentice Hall, Upper Saddle River. ISBN 0-201-56611-7
5. Hughes A (2009) Motors and drives-fundamentals, types and applications. Elsevier. ISBN-13: 978-0-7506-4718-2
6. Mohan N (2003) Electric drives – An integrative approach. MNPERE. ISBN 0-9715292-1-3

7. Krishnan R (2001) Electric motor drives-modelling, analysis and control. Prentice-Hall. ISBN: 0-13-0910147
 8. Heier S (1998) Wind energy conversion systems. Wiley, New York. ISBN: 0 471 97143 X
 9. Petersson A (2003) Modeling and control of doubly-fed induction generators for wind turbines. Licentiate thesis, Department of electric power engineering, Chalmers University of Technology, Göteborg, Sweden
 10. Grauers A (1996) Design of directly driven permanent magnet generators for wind turbines. Doctoral thesis, Chalmers University of Technology, Gothenburg, Sweden
 11. Petersson A et al (2005) Modeling and experimental verification of grid interaction of a DFIG wind turbine. *IEEE Trans Energy Convers* 20(4):878–886
 12. IEC 61000-4-15, Testing and measurement techniques – Flickermeter – Functional and design specifications. Advanced Test Equipment Corp, San Diego, CA
 13. IEC Standard 61400-21 (2001) Wind turbines measurement and assessment of power quality characteristics of grid connected wind turbines, 1st edn. Bureau Central de la IEC/CEI, Genève
 14. Thiringer T, Petru T, Lundberg S (2004) Flicker contribution from wind turbine installations. *IEEE Trans Energy Convers* 19(1):157–163
 15. Thiringer T (1996) Power quality measurements performed on a low-voltage grid equipped with two wind turbines. *IEEE Trans Energy Convers* 11(3):601–606
 16. Lundberg S, Thiringer T (2009) Design considerations for a power electronic converter for series connection of wind turbines. In: Second EPE – Wind energy chapter seminar, April 23–24, 2009, Royal Institute of Technology, Stockholm, Sweden
 17. Thiringer T (2007) Experimental experiences of the summation of flicker contribution from individual wind turbines in a wind park. In: 9th international conference on energy and power systems, IASTED 2007, Clearwater, Florida, USA. January 3–5
- Books and Reviews**
- Abdulahovic T (2009) Analysis of high-frequency electrical transients in offshore wind parks. Thesis, Department of electric power engineering, Chalmers University of Technology, Göteborg, Sweden
- Ackermann T (2005) Wind power in power systems. Wiley, Chichester
- Akhmatov V, Eriksen PB (2007) A large wind power system in almost island operation – A Danish case study. *IEEE Trans Power Syst* 22(3):937–943
- Bollen MHJ, Gu IYH (2006) Signal processing of power quality disturbances. Wiley, Canada
- Bongiorno M, Petersson A (2009) Development of a method for evaluation of wind turbines ability to fulfil Swedish grid codes. Elforsk report 09:25, Elforsk, Frederiksberg
- Burton T et al (2001) Wind energy handbook. Wiley, Chichester
- Cardell JB, Anderson CL (2009) Estimating the system costs of wind power forecast uncertainty. In: IEEE power & energy society general meeting, 2009. PES '09, 26–30 July 2009
- Chen Y, He Y, Bao Y, Shen J (2008) Present situation and future development of wind power in China. In: 3rd IEEE conference on industrial electronics and applications, 2008. ICIEA 2008, 3–5 June 2008
- Dell RM, Rand AJ (2004) Clean energy. Royal Society of Chemistry, Cambridge
- Elkington K (2009) Modelling and control of doubly fed induction generators in power systems. Royal Institute of Technology, Stockholm, Sweden
- Evans J, Shawwash Z (2009) Assessing the benefits of wind power curtailment in a hydro-dominated power system. In: CIGRE/IEEE PES joint symposium on integration of wide-scale renewable resources into the power delivery system, 2009, 29–31 July 2009
- Fadaeinedjad R, Moallem M, Moschopoulos G, Bassan S (2007) Flicker contribution of a wind power plant with single and multiple turbine representations. In: IEEE electrical power conference, 2007. EPC 2007, Canada, 25–26 Oct. 2007
- Fang C (2009) Practical problems and coping strategies of wind power industry in China. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009
- Fang C, Bao C, Qi W (2009) The construction goals and spatial layout project of the offshore three Green Gorges in China. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009
- Fuchs E, Masoum M (2008) Power quality in power systems and electrical machines. Academic Press, Burlington, MA
- Gasch R, Tvele J (2004) Wind power plants: fundamentals, design, construction and operation. Earthscan Publications, London
- Gipe P (2002) Wind power in view wind turbines, aesthetics, and public acceptance. Academic Press, San Diego, CA
- Gipe P (2004) Wind power: renewable energy for home, farm, and business. James & James, UK
- Gu M (2009) Research on strategic importance of non-grid-connected wind power for U.S. energy development. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009
- Gu L, Jin C (2009) Performance evaluation of non-grid-connected wind, power industry cluster. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009
- Hau E (2006) Wind turbines – fundamentals, technologies, application, economics. Springer, New York
- He A, Yan Z, Gu W (2009) Research on non-grid-connected wind power systems used for hydrogen production from water-electrolytic. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009
- Heier S (2009) Grid integration of wind energy conversion systems, Wiley, New York
- Hoffman DL, Molinski TS (2009) How new technology developments will lower wind energy costs. In: CIGRE/IEEE PES Joint

- symposium on integration of wide-scale renewable resources into the power delivery system, 2009, 29–31 July 2009
- Holmgren M, Haarla L, Matilainen J, Holttinen H (2009) Power regulation resources required by wind power in Finland and regulation characteristics of power plants. In: CIGRE/IEEE PES joint symposium on integration of wide-scale renewable resources into the power delivery system, 2009, 29–31 July 2009
- Holttinen H et al (2009) Design and operation of power systems with large amounts of wind power. Final report, IEA WIND Task 25. VTT technical research centre, Finland
- Jin J, Chen Q, Ma Y, Gong C (2009) Non-grid-connected wind power system and its high power DC-DC converter. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, 24–26 Sept 2009
- Kai T, Tanaka A (2008) A new smooth scheme for power fluctuation using inverter of wind power generation with doubly fed induction generator. In: International conference on electrical machines and systems, 2008. ICEMS 2008. Wuhan, China
- Keeley G (2009) Spain's wind turbines supply half of the national power grid. *The Times*, 10 November 2009
- Kundur P (1993) Power system stability and control. McGraw-Hill, New York
- Li M (2009) A study on the exploitation of wind energy and the development of the non-grid-connected wind power industry in central China. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009
- Li M, Fang C (2009) Information system of on-grid and non-grid-connected wind power industry based on SuperMap. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009
- Li DD, Niu LQ (2008) Reliability analysis of electric distribution system integrated with wind power. In: 3rd IEEE conference on industrial electronics and applications, 2008. ICIEA 2008, 3–5 June 2008
- Ming L, Xiong F, Yue W, Leqiang Z, Zhaoan W, Gang L, Weizheng Y (2009) Overvoltage protection of high power IGBTs in wind power converters under short circuit. In: IEEE 6th international power electronics and motion control conference, 2009. IPEMC '09, 17–20 May 2009
- Lin HC (2006) On line web-based maximum wind power monitoring and control system. In: IEEE international conference on systems, man and cybernetics, 2006. SMC '06, vol 5, 8–11 Oct 2006
- Liu J, Schweizer D, Seiler K (2009) Wind power interconnection and integration roadmap: An RTO's perspective. In: IEEE power & energy society general meeting, 2009. PES '09, Calgary, 26–30 July 2009
- Lobo MG, Sanchez I (2009) Aggregated wind power prediction methods based on distances between weather forecasting vectors. In: International conference on clean electrical power, 2009, 9–11 June 2009
- Luickx PJ, Delarue ED, D'haeseleer WD (2009) Effect of the generation mix on wind power introduction. *IET Renew Power Gener* 3(3):267–278
- Manwell et al (2002) Wind energy explained – theory, design and application. Wiley, Chichester (Recent years have seen a growth in the implementation and economic viability of wind energy technology. This safe and abundant source of clean, renewable energy is now making a significant contribution to electricity supplies worldwide. Addressing the growing requirement for information on the theory and practical application of wind technology, Wind Energy Explained provides a through introduction to this multi-disciplinary field)
- Max L (2009) Design and control of a DC collection grid for a wind farm. Thesis, Department of electric power engineering, Chalmers University of Technology, Göteborg, Sweden
- Meei-Song K (2007) Generation cost assessment of an isolated power system with a fuzzy wind power generation model. *IEEE Trans Energy Conversion* 22(2):397–404
- Meibom P, Weber C, Barth R, Brand H (2009) Operational costs induced by fluctuating wind power production in Germany and Scandinavia. *IET Renew Power Gener* 3(1):75–83
- Meier S (2009) System aspects and modulation strategies of an HVDC-based converter system for wind farms. Royal Institute of Technology, Stockholm, Sweden
- Mohan N et al (2003) Power electronics – converters, applications and design. Wiley, New York
- Muljadi E, Butterfield CP, Chacon J, Romanowitz H (2006) Power quality aspects in a wind power plant. In: IEEE power engineering society general meeting, 2006, Montreal, Canada
- Muljadi E, Mills Z, Foster R, Conto J, Ellis A (2008) FAULT analysis at a wind power plant for one year of observation. In: IEEE power and energy society general meeting – conversion and delivery of electrical energy in the 21st century, 2008, 20–24 July 2008
- Muljadi E, Nguyen TB, Pai MA (2009) Transient stability of the grid with a wind power plant. In: IEEE/PES power systems conference and exposition, 2009. PSCE '09, Seattle, 15–18 March 2009
- Obersteiner C, Sague M (2009) On the market value of wind power. In: 6th international conference on the European energy market, 2009. EEM 2009, 27–29 May 2009
- Paensuwan N, Yokoyama A (2009) Risk-based dynamic TTC calculation in a deregulated power system with a large penetration of wind power generation. In: CIGRE/IEEE PES joint symposium on integration of wide-scale renewable resources into the power delivery system, 2009, 29–31 July 2009
- Ravn HF (2006) Short term wind power prognosis with different success criteria. In: International conference on probabilistic methods applied to power systems, 2006. PMAPS 2006, 11–15 June 2006
- Shinde SM, Patil KD, Gandhare WZ (2009) Dynamic compensation of reactive power for integration of wind power in a weak distribution network. In: 2009 international conference on control, automation, communication and energy conservation, 2009. INCACEC 2009, 4–6 June 2009
- Tanabe T, Sato T, Tanikawa R, Aoki I, Funabashi T, Yokoyama R (2008) Generation scheduling for wind power generation by

storage battery system and meteorological forecast. In: IEEE power and energy society general meeting – conversion and delivery of electrical energy in the 21st century, 2008, 20–24 July 2008

Thakur T, Goyal S, Gambhir J, Kaur I (2008) Optimisation of imbalance cost for wind power marketability using hydrogen storage. In: Joint international conference on power system technology and IEEE power India conference, 2008. POWERCON 2008, 12–15 Oct 2008

Tsikalakis AG, Katsigiannis YA, Georgilakis PS, Hatzigiorgiou ND (2006) Determining and exploiting the distribution function of wind power forecasting error for the economic operation of autonomous power systems. In: IEEE power engineering society general meeting, 2006

Ummels BC, Pelgrum E, Kling WL (2008) Integration of large-scale wind power and use of energy storage in the Netherlands' electricity supply. IET Renew Power Gener 2(1):34–36

Wämundson M, Hassan F (2009) Optimum DC-link solution in HVDC wind park actively interfaced to the grid. Elforsk report 08:45, Elforsk, Frederiksberg

Wang J, Shahidehpour M, Zuyi Li (2009) Security-constrained unit commitment with volatile wind power generation. In: IEEE power & energy society general meeting, 2009. PES '09, 26–30 July 2009

Wang L, Wei J, Wang X, Zhang X (2009) Wind power interconnection and integration roadmap: An RTO's perspective. In: IEEE power & energy society general meeting, 2009. PES '09, 26–30 July 2009

Wang L, Wei J, Wang X, Zhang X (2009) The development and prospect of offshore wind power technology in the world. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009

Wang L, Dong L, Hao Y, Liao X (2009) Wind power prediction using wavelet transform and chaotic characteristics. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009

Watson R (2001) Large scale integration of wind power in an island utility—an assessment of the likely variability of wind power production in Ireland. In: IEEE porto power tech proceedings, 2001, vol 4. Portugal

Xing-wei X, Gang M, Guang-hui S, Hong-peng Z, Kai-yuan H, De-bin G, Jia-qi T, Xin M, Yong X (2009) The problems and solutions for large-scale concentrated integration of wind power to partially weak regional power grid. In: International conference on sustainable power generation and supply, 2009. SUPERGEN '09, 6–7 April 2009

Yan Z, Gu W (2009) Research on integrated process of large-scale non-grid-connected wind power and coal-to-ethylene glycol. In: World non-grid-connected wind power and energy conference, 2009. WNWEC, Nanjing, 24–26 Sep 2009

Zha L, Song J, Liu H (2007) Simulation study of wind power with continuously variable transmission. In: 2nd IEEE conference on industrial electronics and applications, 2007. ICIEA 2007, 23–25 May 2007

Wind Power Grid Integration: Transmission Planning

DALE OSBORN

Midwest ISO, Columbus, OH, USA

Article Outline

Adjustment of Economic Simulation Model Transmission Line Ratings

Economic Studies

Value-Based Long Range Planning Method

Wind Integration Studies

Other Information to Answer Questions

Conclusion

Bibliography

How is transmission planning done? In as many ways as there are different goals of many organizations influenced by regulation. The discussion will relate to entities connected to a large interconnected grid.

Three types of organizations of utilities exist in the USA:

1. The utilities who participate in a Regional Transmission Organization (RTO). RTOs may consist of utilities from several states or one state as in the case of the New York Independent System Operator or California System Operator. RTOs have planning processes that incorporate the entire footprint of the RTO. Most of the RTOs operate markets.
2. Utilities who do not participate in a RTO. The individual utilities have their own planning processes. Some of the utilities have relationships with other utilities for joint planning.
3. Merchant transmission that plans transmission and seeks participants to help fund the transmission project.

The National Electric Reliability Corporation (NERC) has the responsibility of enforcing power system reliability operating and planning standards in the USA and coordinating with Canada. Every utility in the USA and Canada participates in the NERC reliability assessments to assure that the transmission system will meet standards and is expected to perform reliably.

Most of the criteria for planning of transmission are based on the NERC standards.

There are three large areas or interconnections that operate as interconnected systems in the lower 48 states of the USA. The three interconnections are weakly tied with HVDC back to back transmission. Hydro Quebec is also tied to the USA and Canada with HVDC ties. ERCOT (TEXAS) is an interconnection on its own. Alaska has its own systems.

Transmission is planned to serve the native load plus purchases and sales transactions. Models are built on the interconnection level so that the interaction of all native load and transactions in the USA and Canada can be studied simultaneously with a major transmission line model. NERC coordinates the model building process.

The transmission models are obtained by a utility or an RTO and then the detailed transmission in their areas and sometimes sub-transmission (below 100 kV) are added. Thus, any utility or RTO has a detailed model of their footprint and a good representation of how they will react with the transmission system adjacent to them.

Energy markets, larger footprints of the RTOs compared to utilities, the insertion of significant quantities of wind and solar generation, and the federal government funding of coordinated planning at the interconnection level has recently initiated integrated studies on a much larger scale than has previously been done in the USA. One primary driver for the studies is to plan transmission for the use of wind, solar, and other renewable resources. Wind, solar, and other renewable resources are not located near the major loads. Transmission will have to be constructed to take advantage of renewable resources. The transmission needed for renewable resource is for energy that provides an optional fuel source and is not necessarily needed for reliability purposes. Economic criteria for building transmission are entering the transmission planning processes.

The present US transmission system is similar to a net with weak points between the geographic boundaries of organizational and former organization zones. State boundaries, country boundaries, RTO boundaries, former utility boundaries, and geographic restrictions such as lakes and national parks are usually the location of the weak links in the net that create

holes. There are strong ropes that reinforce the net in some areas.

There are three basic philosophies concerning transmission design:

1. Do only what is needed. This is the present reliability process for transmission planning. This has worked for years, but does not address the characteristics and economics of renewable energy very well.
2. Plan for reliability and reinforce the system at the constraint locations. This looks like a good approach until one analyzes the large power transfers that renewable energy resources, such as wind energy, would require for efficient energy markets. Rebuilding all the AC underlying system to transfer high power transfers and the control of loop flow is a formidable task. There are over 300 constraints in the EWITS [1] study. Fixing one constraint does not mean that the constraint just moves over to another area. The area that the constraint moved would then need to be fixed also with a transmission expansion. Separating which fixes are needed for reliability and how much of the fix is related to the market is a quagmire that may not be possible to be resolved.
3. Design a transmission overlay that would produce an efficient system of markets including the effects of renewable energy without impacting the underlying AC system significantly. Reliability would be mainly addressed in the transmission planning of the underlying AC system. The overlay would be designed to be self-contingent, economical, and not create reliability problems either in the overlay or the underlying system. One advantage of the overlay philosophy is that the transmission design and the cost of the overlay can be separately identified and not comingled with the reliability upgrades. The transmission planning process to make the markets efficient can be addressed separately from the existing transmission planning processes that design the underlying AC system.

The most likely choice of the three philosophies is that all of the above will be used. A transmission system design with a top layer overlay that delivers bulk energy to the energy market, a second layer that consists of RTOs and large utilities, and a third layer of utilities not in a RTO, public power, and possibly

federal power agencies may be envisioned for the Eastern Interconnection. ERCOT may or may not be further interconnected with HVDCties. The Western Interconnection may have another organization based on coordination without energy markets.

The following is a simplistic explanation of the levels or layers of transmission planning.

Municipals and small utilities probably can do most of their planning with power flow studies to check for overloaded lines and to maintain voltages within acceptable limits. Contingency studies are run to confirm that the system will not be adversely affected for probable events. Generally the NERC planning standards are used as criteria. Short circuit and relay coordination studies are done to assure that the equipment is within ratings and properly protected for faults. The outside transmission system adequacy is usually studied by a NERC region or a RTO in their area. The engineers may participate in the NERC regional studies or the RTO meetings for coordination of plans. Reliability is the focus of the studies. Engineers at this level are very specific about the details of the design of the transmission system. Economic and technical evaluation of multiple alternatives is usually performed for each prospective project. Larger utilities may have a division that studies individual areas on their system in a similar manner. The NERC region or the RTO usually does not study their system to the detail that they do.

Larger utilities have complete planning staffs with specialists in areas such as dynamics. They perform their own studies for the generation interconnection, transmission expansion, dynamics, possibly equipment specification transient studies and transmission service requests.

Providing testimony to support proposed plans is a transmission planning function.

Very large utilities may break the transmission system into voltages or regions and have multiple sections of transmission planners. These utilities may have a research department.

There is usually a separate department for protection studies and short circuit studies. There are usually separate departments for generation expansion, production cost simulations, market use studies, resource adequacy, and load forecasting outside the planning department.

If a larger utility is in an RTO, the generation interconnection studies may be performed by the RTO or under contract with them. The larger utility would probably participate in RTO planning meetings and on task forces. Regional transmission expansion studies involving multiple utilities are probably performed by the RTO. The larger utilities usually run their own version of the RTO studies for verification of results and to test their own concepts.

The processes for expansion planning vary greatly from RTO to RTO.

Some of the common processes are as follows:

1. Studies to determine the ability of a generator to deliver to a utility load or to the RTO as a whole.
2. Generation Interconnection and expansion studies.
3. Longer-term, larger power transfer transmission service studies to shorter-term transmission service studies are approved or denied by a tariff administrator using a program to evaluate the current status service available. There are too many transmission service requests for a study to be done for each request and the time is not sufficient many times to do a study.
4. Available Transmission Capacity (ATC) studies are the link from planning to operations and tariff administration for transmission service requests. ATC is published multiple times per day.
5. Shorter-term transmission expansion planning up to 10 years in advance. The transmission that is approved for construction comes from these plans. These are bottom-up plans. Bottom-up planning starts with the system today and sequentially determines the best set of alternatives to provide reliable service in the future. Power flow, power transfer sensitivity, voltage stability, dynamics, economic evaluation of the alternatives, small signal stability, and large amounts of contingency analysis are studies performed annually. The plan is approved by the Board of Directors for construction by the Transmission Owners. RTOs do not own transmission.
6. Constraint mitigation is also evaluated by the RTOs.

The part where the RTO planning methods depart from each other and large utilities is in the long range transmission expansion planning.

Most of the RTOs use a multi-scenario AC study analysis followed by an economic analysis in a bottom-up process to determine the long range transmission expansion plans. The methods used are very similar to their short-term transmission expansion plans.

The Midwest ISO uses a bottom-up short-term transmission expansion process similar to the other RTOs. A mid-range Midwest ISO plan that will accommodate the present 23,000 MW of Renewable Portfolio Standard requirements (wind) for the Midwest ISO footprint is being developed to specify the sequence of transmission expansion from the short-term to the long-term expansion plan. The long-term expansion plan would be some future expansion similar to the EWITS or the EIPC [2].

The Midwest ISO uses a top-down value-based approach with an end year of 15 years out for the long range plan. Multiple scenarios or futures are studied that determine the probable outcomes of the conditions that the power system may be expanded. Economic simulations are run to provide economic and power flow performance data from a simulated energy market for 8,760 h in a year. Generation forecasts are optimized for each future. Differences between the economic simulations produce rough budget information that can be used to determine if and what transmission might be considered. Transmission plans are developed from the differences of economic and power flow information. Graphical outputs and tables are used to determine the location of conceptual transmission expansion and their size.

Multi-RTO and large utility studies on for the Eastern Interconnection have been made completed. A study for the Western USA is expected this year. ERCOT does its own studies. Canada is doing its own Pan-Canadian study. DOE/NREL is sponsoring an all US Renewable Energy Future Study for the year 2050 for up to 80% renewable generation of all types which should be out later in 2010. DOE funded coordinated transmission studies for the Eastern Interconnection and the Western Interconnection that are just starting.

One major limitation to coordinated studies is time-synchronized data for renewable generation. The Eastern Wind Integration Study produced a publicly available data set for each 10 min for the parts of the Eastern Interconnection having wind. Having time-synchronized data enables the value of geographic

diversity and the availability of wind generation to serve the peak load, capacity credit. Time-synchronized data allows the location of wind resources in different areas such as onshore and offshore. The Western Wind and Solar Integration Study also has wind data similar to the EWITS wind, but the two databases are not time-synchronized due to different vendors producing the data. Solar data lags wind data in development, but progress is being made. The data from Canada is not synchronized with the US data. Political borders affect data generation as well as transmission planning.

Figure 1 shows the capacity factors of wind units for the eastern USA in the area covered by the EWITS database for wind production.

Figure 2 shows the output from the EWITS wind output database for the maximum wind output for the years. Red and orange colors show high wind energy production. Blue represents poor production. Some areas that traditionally have excellent wind outputs are not on peak.

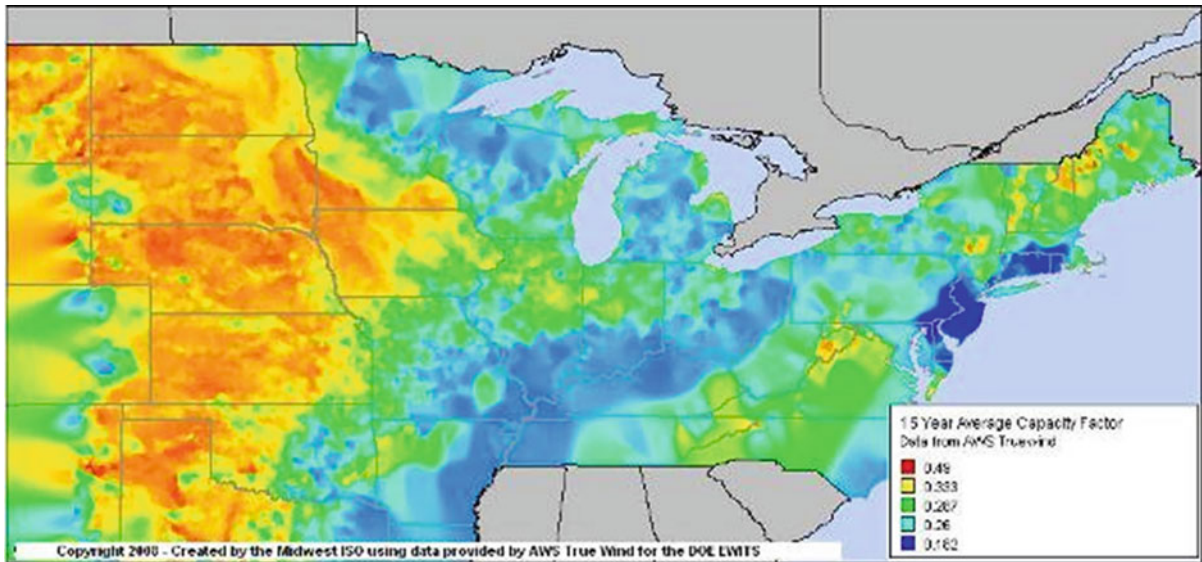
Figure 3 shows the worst day of a year. Almost all the wind over the eastern USA is off. Arguments about wind not being a reliable resource often refer to plots like Fig. 3. The focus of studies is to find out what can be done with the wind resource. The overall system is designed to fill the gaps left by wind energy to produce a reliable and economical system.

The diversity of wind is greater in the north-south direction in the eastern USA than for the west-east direction. The plot in Fig. 4 shows that long distances are required to reduce the correlation of wind energy production.

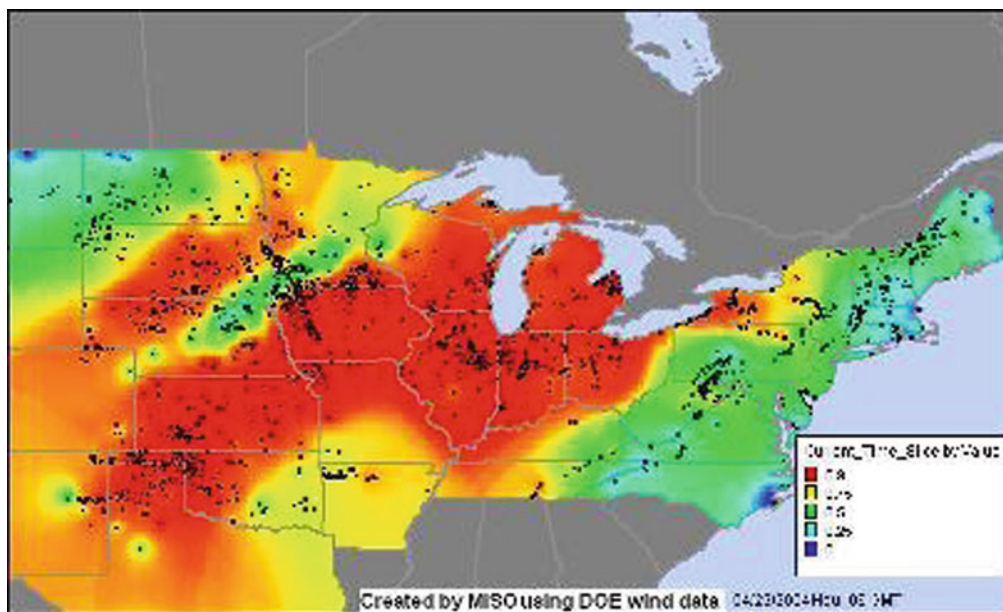
Figure 5 shows the reduction in variability from the EWITS study as the wind resource is expanded over the eastern USA and tied together with transmission.

The variation of wind is reduced by a factor of 4. Transmission can link the geographically diverse regions of the eastern USA together to produce a better energy product that would require fewer reserves to manage. Note there is a level of saturation as the wind resource exceeds 5,000 MW.

Not only does wind vary from minute to minute, the yearly output can vary by up to 26% from the maximum output year to the minimum output year. A study by WindLogics reported this result for an examination of a 12-year time span using forecasted information. Variations of this magnitude produce



Wind Power Grid Integration: Transmission Planning. Figure 1
Capacity factors of wind units for the eastern USA

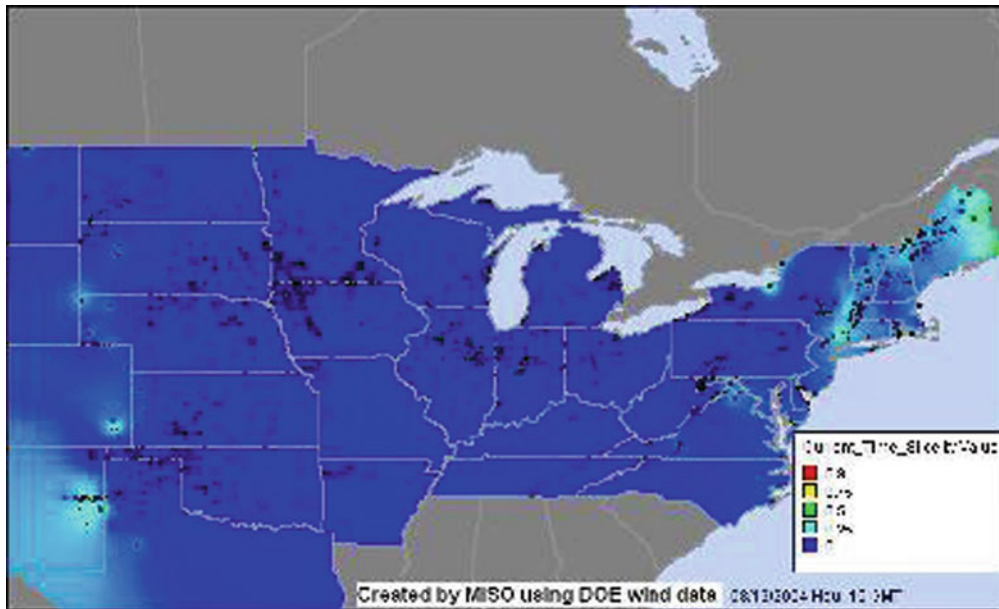


Wind Power Grid Integration: Transmission Planning. Figure 2
Maximum wind output for the years. Red and orange colors show high wind energy production. Blue represents poor production

different generation patterns that must be included in the design of transmission systems. Multiple years of study data are run through the economic simulations to determine annual variability patterns.

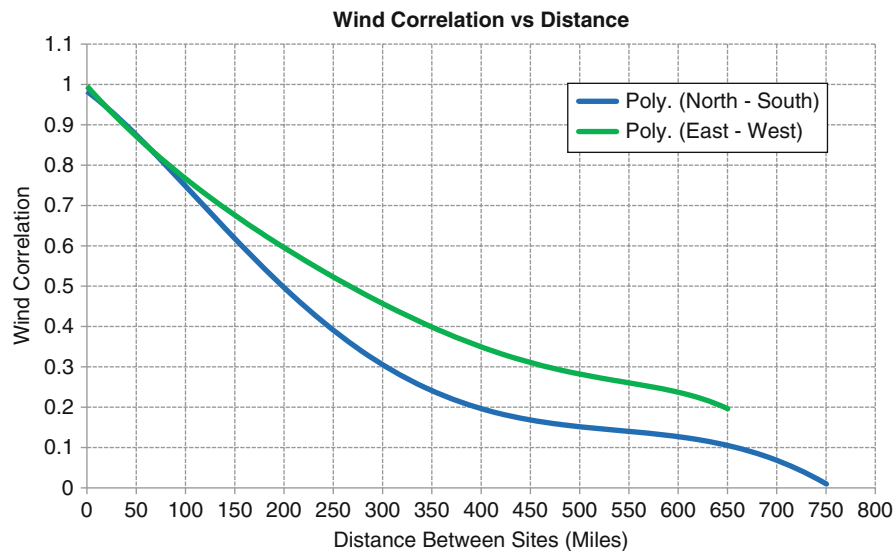
Unfortunately, only 3 years of hourly data exist for the EWITS database.

The aggregation of wind energy also increases the probability that the wind generation would be available



Wind Power Grid Integration: Transmission Planning. Figure 3

Worst day of a year. Almost all the wind over the eastern USA is off



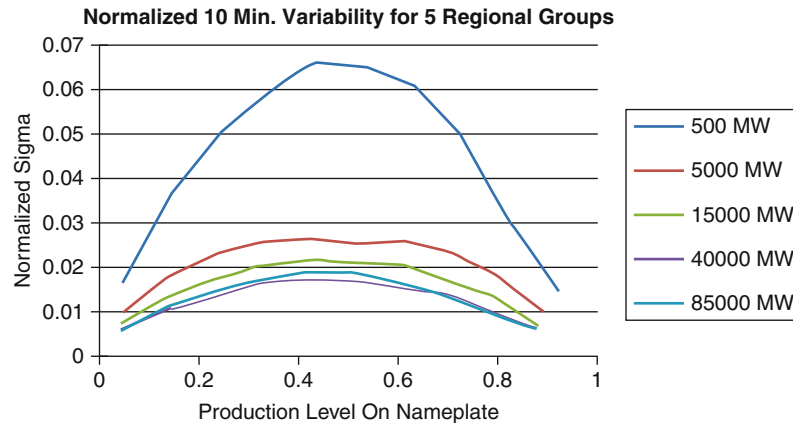
Wind Power Grid Integration: Transmission Planning. Figure 4

Long distances are required to reduce the correlation of wind energy production

on peak. A loss of load probability expectation calculation is performed that shows the component due to aggregation of resources and the component due to the transmission overlay for 3 years. The transmission overlay has a large impact when the wind is located

primarily in the western (Scenario 1) part of the Eastern Interconnection than in the eastern part (Scenario 3).

The EWITS study, Fig. 6, shows that the capacity credit of wind for a large aggregated area can be in the range of 24%. This is three times higher than the

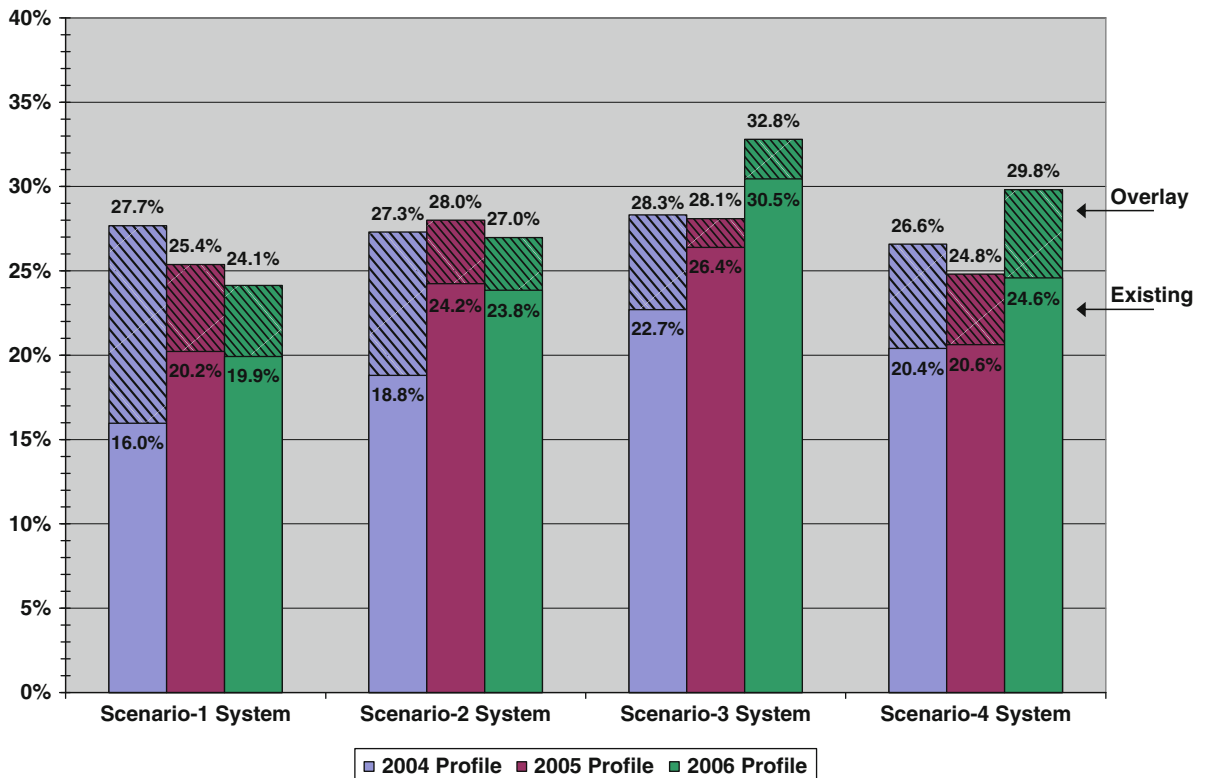


Wind Power Grid Integration: Transmission Planning. Figure 5

Reduction in variability from the EWITS study as the wind resource is expanded over the eastern USA and tied together with transmission

Study System ELCC Scenarios (1 - 4)

Existing & Overlay Transmission Tie Limits - ELCC (%) (Shaded Area shows Increased ELCC of Overlay)



Wind Power Grid Integration: Transmission Planning. Figure 6

The EWITS study shows that the capacity credit of wind for a large aggregated area can be in the range of 24%

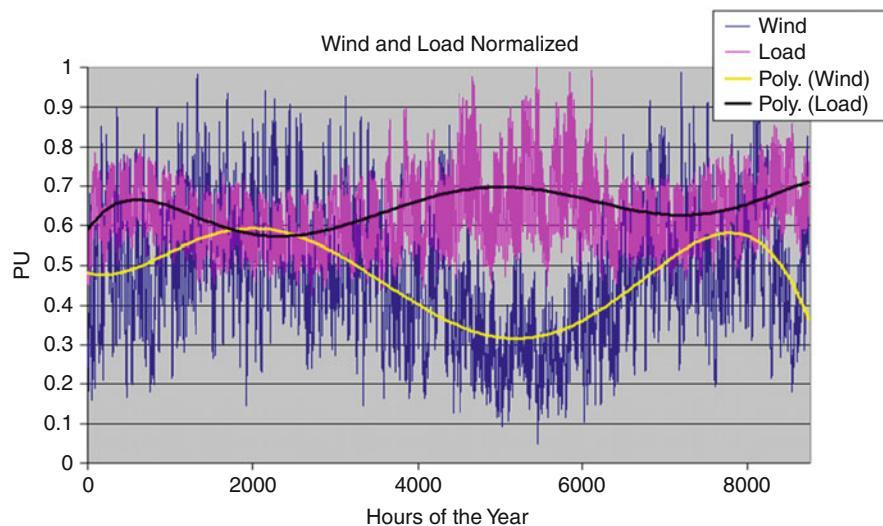
present Midwest ISO wind capacity credit of 8%. Less conventional generation is needed for the generation forecast with a higher wind capacity credit. The amount of wind generation is not reduced as the wind generation resource has to be large enough to produce the RPS requirements. Transmission helps the process, but the geographic diversity of the wind resource is the primary factor in improving the wind capacity credit.

Figure 7 is a representation of wind energy production hourly for a year normalized to the output of load in the area. Curve fits through the highly variable plots enable one to see trends. The load curve has a summer peak with minimum load during the spring and fall. The wind has a minimum output during the peak period for the load and maximum outputs during the spring and fall when the load least needs generation. Wind is also more variable than the load during the spring and fall periods.

If the load and the corresponding wind energy output are sorted in descending order by load, a load duration curve is produced which shows the amount of time in which the load level is achieved (Fig. 8). The peak load occurs only for 1 h of the year. The trend curves show that wind generation has an average base load characteristic off-peak and a reduced capability on

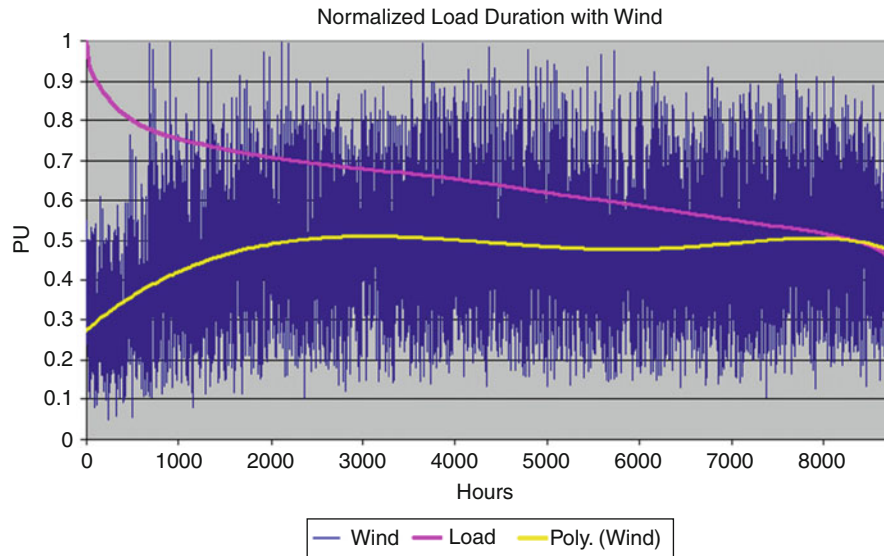
peak. Thus, the wind generation would offset base load generation production during the off-peak periods. Base load generation needs to have about a 60% capacity factor (average output divided by maximum output over the year). Wind decreases the base load capacity factor to the point that a combustion turbine or low capacity factor combined cycle plant may be the best selection for the generation forecast when combined with wind energy.

The mix of generation types to serve the load is a combination of the capital costs to construct a generator and the energy costs to operate the generator. Base load generation has higher capital costs and lower fuel costs. Gas fired generation has lower capital costs and higher fuel costs. Generation forecasting programs are run to determine the optimal mix of generation that is needed to reliably serve the load. When operating generation with low costs operating costs are run first. The higher cost operating cost generation is run next until the generation is stacked in merit order. Figure 9 shows a simplified generation merit order stack. The area under the load duration curve to the left is the energy supplied for the load. The area above the load duration curve and to the right is energy that may be bid into the energy market for sale. The area between the purple line and the yellow line depict the



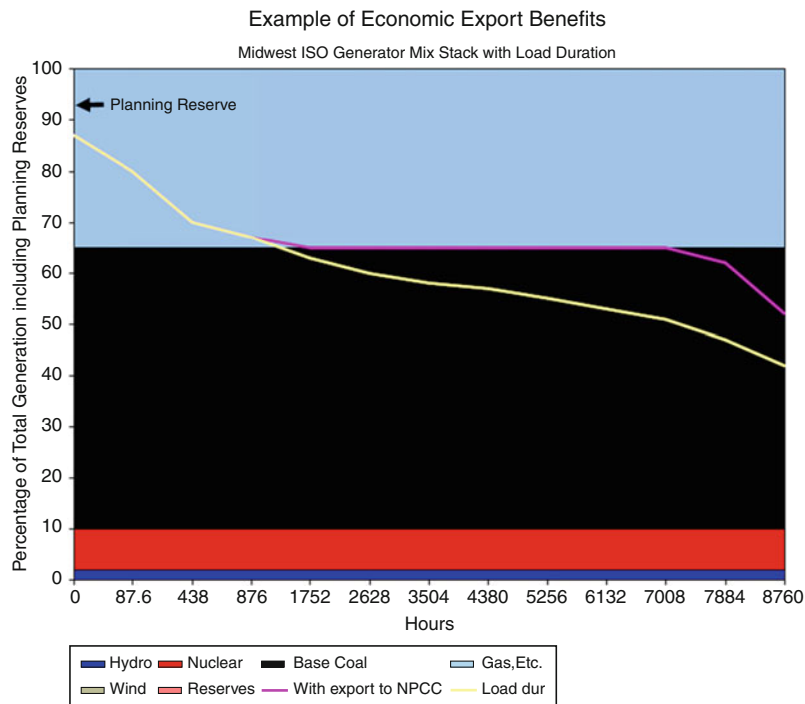
Wind Power Grid Integration: Transmission Planning. Figure 7

Hourly wind energy production for a year normalized to the output of load in the area



Wind Power Grid Integration: Transmission Planning. Figure 8

Load duration curve (*magenta*), which shows the amount of time in which the load level is achieved. The peak load occurs only for 1 h of the year



Wind Power Grid Integration: Transmission Planning. Figure 9

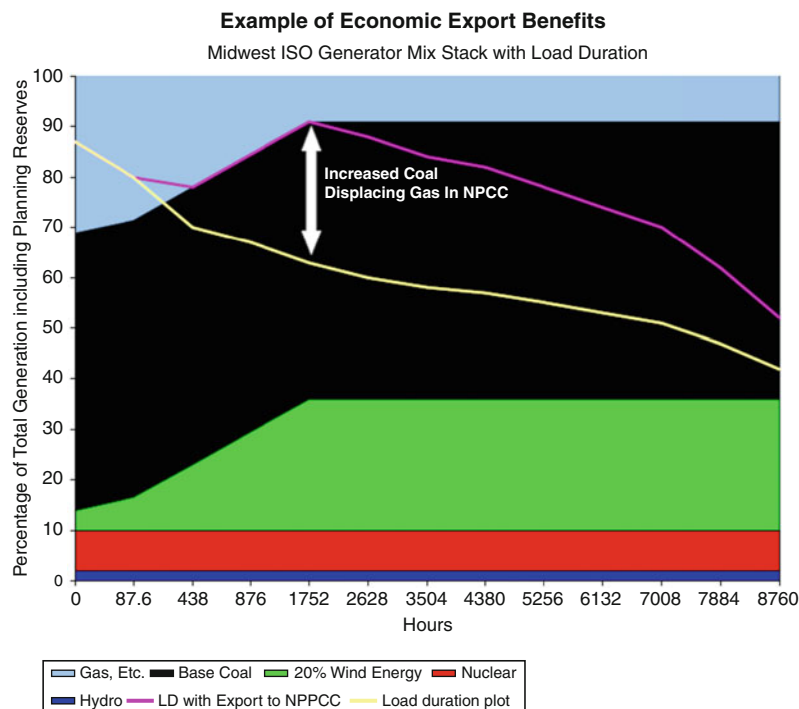
All generation is used to meet the peak load. Generation that can be used to meet load has a 100% capacity credit for this example

energy that might be sold on the energy market. The purchasing party has limitations on how much energy it can accept.

All generation shown in Fig. 9 is used to meet the peak load. Generation that can be used to meet load has a 100% capacity credit for this example.

Figure 10 shows the result of injecting wind generation into the generation mix. Coal generation is forced upward on the stack. Coal is forced into an area usually occupied by gas generation in an optimal generation mix. The capacity factor of coal is reduced. More coal generation is available for sale. The area between the purple line and the yellow line is energy that might be sold on the energy market. The purchasing area cannot receive all the available energy. The coal generator must be reduced in output to match the load plus purchase chart. Even with sales, the coal capacity factor is reduced compared to Fig. 9, and this is a benefit for global warming. For the purchaser, the use of coal energy may be against their green commitment. Is the goal global warming or local contribution to global warming?

For this example wind has a capacity credit of about 5%. If peak load and energy grow at the same rate, the vertical axis will increase and the graph will be proportional to Fig. 10. If the average load factor (load energy divided by the peak load times 8,760 h per year) is 60% and the average wind generator capacity factor is 30%, then 2 MW of wind generation has to be added for each MWH of energy growth. If the RPS wind growth requirement is 20% of the load energy, then the wind generation to be added would be 0.4 MW. Assume the generator reserves are about 12%. Reserves provide some extra generation if something would happen to a generator at the peak period to replace the generator with the outage. Gas generation would be added for 1.10 MW (1.12 minus 2% credit for the wind). For each MW of load added, 1.5 MW of gas and wind generation would be added. The mix of coal would be very gradually reduced. At the present rates of load growth it would require 15 years to adjust the generation mix to an economically optimal mix.



Wind Power Grid Integration: Transmission Planning. Figure 10
Example of economic benefits

Adjustment of Economic Simulation Model Transmission Line Ratings

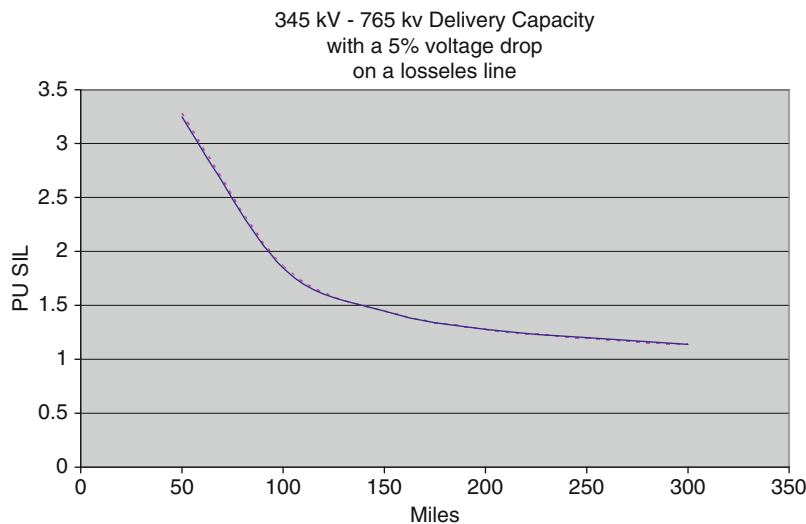
Economic models use linear or “DC” representation of transmission lines that do not have distance considerations considered. A linear line is modeled as having a MW rating and an impedance. There are no voltage drop, dynamic rating considerations, and distance considerations. AC power flows and dynamic studies automatically consider all the factors due to the mathematical representation of the lines. The power flow models used by economic simulation programs use power flow data that may have a thermal rating that is higher than what can actually be carried on the transmission line. To resolve this issue, the economic simulation models have lines rated according the St. Clair curve shown in Fig. 11 that accounts for the decreasing power transfer capability of a transmission line with distance. Lines in the power flow case that is an input to the economic simulation case have ratings that are adjusted by the length of the transmission line and the voltage of the line. For 765 kV a 1.2 times multiplier to the St. Clair curve is used. For a 345 kV line a 1.5 times multiplier is used. The multiplier is chosen to reflect power flow values that are achieved during contingency operation. Adjusting the ratings of transmission lines produces more realistic line loadings for the economic simulation than that produced using the

ratings from a power flow case. The surge impedance loading (SIL) of a 345 kV line is approximately 400 MW. The surge impedance loading a 765 kV line is about 2,260 MW. The rating of an 800 kV HVDC line is about 6,400 MW.

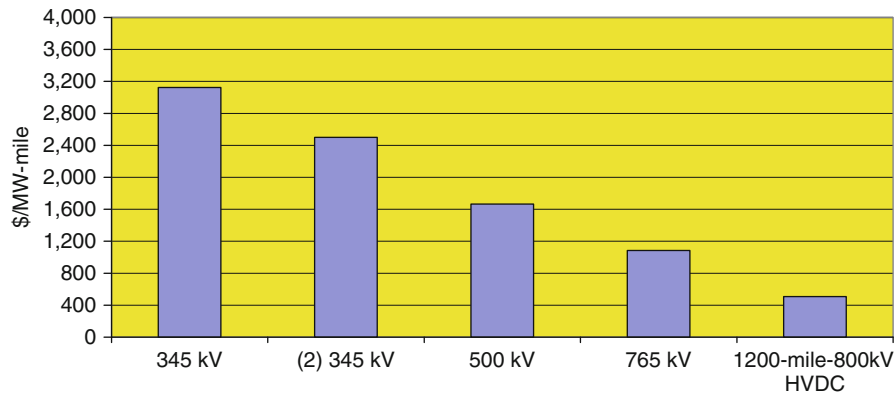
Figure 12 shows the relative cost of constructing a transmission line to transmit 1 MW 1 mile. The 765 kV and HVDC appear to be the best options. However, the line choice has to consider distance and loading.

AC versus HVDC is also an option that must be considered in transmission overlay design. HVDC lines cost less than overhead AC lines of the same voltage. HVDC convertor stations cost more than AC substations. Figure 13 shows the distance to consider an 800 kV HVDC line rather than a 765 kV line to is about 600 miles. The overlays for JCSP [3] and EWITS used this guideline for the two and three terminal HVDC lines included in the design.

The HVDC lines need to be loaded to be chosen for the overlay design. The steps on the plot in Fig. 14 represent a transmission line of either AC or HVDC. The plot shows that at least two 800 kV lines with a power transfer of 4,300 MW are needed before HVDC is considered to be an option. If the AC system had a 1,500 MW contingency level, the loading level on three HVDC lines to form a self-contingent system for

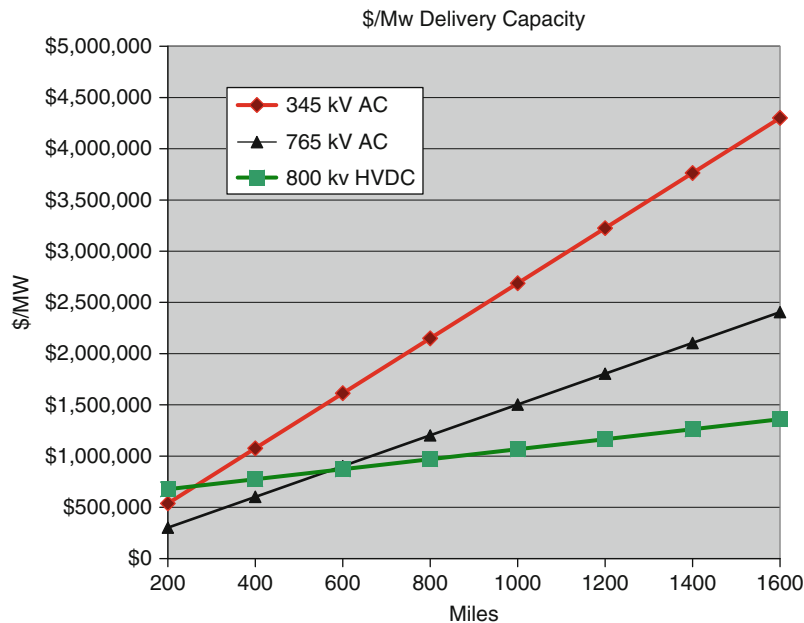


Wind Power Grid Integration: Transmission Planning. Figure 11
Decreasing power transfer capability of a transmission line with distance



Wind Power Grid Integration: Transmission Planning. Figure 12

Relative cost of constructing a transmission line to transmit 1 MW 1 mile

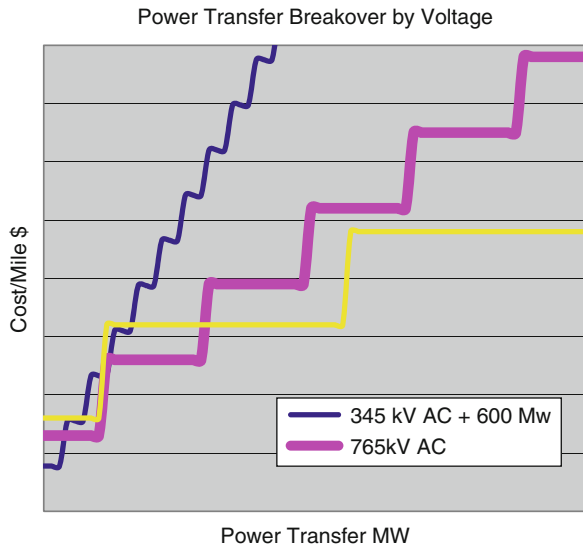


Wind Power Grid Integration: Transmission Planning. Figure 13

Transmission line cost as a function of distance for three voltage levels

the loss of one HVDC line would be 7,900. The HVDC maximum schedule per line would be about 3,950 MW. This may work and have to be tested, but the value is below the threshold. For a three-line HVDC system the HVDC line schedule would be 14,300 MW. Each HVDC line would be scheduled at 4,766 MW, which is above the threshold. The economic simulations for the JCSP and the EWITS showed that this loading

would occur about 80% of the time on the HVDC lines. This roughly validates the Rule of Three. The HVDC overlay would carry only energy and not capacity. This allows the overlay to be loaded heavier than if it had to deliver the full power schedule after a contingency under a security constrained criteria. The generation forecast for each region has enough generation to serve the load. Each area is not dependent



Wind Power Grid Integration: Transmission Planning.

Figure 14

Power transfer breakover by voltage

upon imported capacity. The HVDC lines would run at rated load for no more than 30 min. With a 5-min dispatch interval in the energy market, it is highly possible that the security constrained schedule would be reached in 5–10 min.

A recent short investigation into underground HVDC delivery includes:

1. Superconducting plus and minus 200 kV HVDC at about 10,000 MW with a spare pole to create a self-contingent system using Voltage Source Converters of about 750 MW each. The higher converter count would be able to link more market segments into one HVDC line. The spare pole decreases the number of lines needed to construct a self-contingent system.
2. Gas insulated transmission has the capacity to deliver 800 kV HVDC with a similar spare pole configuration or a metallic return for a self-contingent system. The fewer number of larger converters maybe a problem to locate. While gas insulated transmission has been used for years in some form, the long distance aspects have not been addressed.
3. Electric pipe is a XPLE nanoparticle insulated heavy conductor in a pipe. The increased cross-sectional

area of the conductor reduces the heat losses to values that can be dissipated through the pipe. Higher currents than cable are possible.

While HVDC breakers and networks have been discussed with the underground options, it may be possible to link separate HVDC lines with AC backup transmission to transfer power to healthy poles similar to the overhead HVDC design used in the JCSP and EWITS. The AC backup design may be possible to have the initial HVDC lines constructed without having to wait for the development, testing, design modifications, and retesting of HVDC breakers and networks.

The high level of activity with the submarine HVDC cables indicates that larger HVDC links may be possible. A constrained and a “copper sheet” difference case may provide insight into the scope of the problem. The difficulty is the construction of a European model for the production cost simulation including transmission such as that used in PROMOD [4], GE MAPS [5], or GRIDVIEW (ABB) [6].

Economic Studies

There are many questions that need to be answered when considering a transmission expansion study including wind. Most of the questions focus on economic impacts, carbon reduction, impact on other generation, wind curtailment, and total cost of different futures. There are few questions about the voltage levels or the power flow on transmission lines. Many different factions are involved in the political and economic processes influencing the design of the transmission system. The studies need to be able to address the questions.

Economic drivers that create value from having transmission are as follows:

1. The energy markets in the Eastern Interconnection of the USA and elsewhere have areas with lower cost resources and loads with higher priced energy. Connecting a lower cost resource to a higher priced load area with a transmission line displaces the local high-priced generation with the low-cost resource and lowers the price to the load. The low-cost generation resource may receive more revenue than it would have without the transmission. HVDC transmission produces

higher revenue for the generating resource compared to an AC transmission system. Revenue from an HVDC overlay is not distributed among the intervening generation and diluted by transmission constraints as is the case with AC transmission designs.

2. Wind generation creates large surpluses of energy production capability during the off-peak period. Wind generation has a very low marginal cost. Wind pushes coal generation upward in the merit order stack of generation so that less energy is by coal for the local use that took the wind energy. The coal energy is considerably lower priced than natural gas fired and oil fired generation. If transmission capacity is available, the coal energy replaces the natural gas energy at the higher priced area. The capacity factor of wind generation in good wind areas can be in the 40% capacity factor range. The gas generation displaced by coal has lower capacity factors. Natural gas produces about 60% of the carbon dioxide as coal per MWH. The carbon savings based on fuel usage is not enough to offset the difference in capacity factor reductions. Thus, the total carbon dioxide production is reduced (see JCSP example).
3. Generation capital and fixed operation costs are paid for by the utility or RTO purchasing the generation capacity. Only the marginal cost of energy (fuel and variable operation and maintenance costs) is used to determine the cost of the energy in a market. With wind energy there is no fuel cost, and the operational and maintenance cost is small. If wind energy is placed into the energy market it is usually bid as a price taker. This happens when wind is to be curtailed because of minimum generation requirements. Coal generation is difficult to start after a shutdown. Therefore, coal generation has to be left on line at minimum generation levels to meet the next day's load. If the load is less than the sum of all the minimum generation in an area, wind may have to be curtailed. A price taker has a very good chance of being a successful bidder in an energy market as this is a bottom price. A price taker receives the clearing price of the energy market. Gas generation usually sets the clearing price on the east coast of the USA. For a natural gas price of \$8/MBTU the cost of energy from a combined cycle may be in the \$60/MWH range. Coal would be in the \$20–\$40/MWH range depending on the location. Wind would have a cost in the \$4/MWH range. Thus, if the market clearing price was based on natural gas generation and there was sufficient transmission to deliver the wind energy, the wind generator revenue would be \$60/MWH minus losses and transmission costs. Assume the cost of transmission and losses is about \$50/MWH. Thus, the wind generator would have a net income of \$6/MWH. The alternative may be curtailment of generator output. Transmission provides larger markets for wind energy that allow the full output of the wind generators.
4. Transmission decreases the variability of wind by aggregating the variation of wind generation over a large geographic area by a factor of about 4. Having less variation of the wind generation reduces the reserve capacity that must be online and spinning or held as a fast start reserve generator if the wind would quit.
5. Onshore wind generation has a poor contribution to the peak capacity of a system. The lowest output of the wind generation in a year is the summer season. The wind primarily is present at night and not during the peak period of the day. The capacity credit for the Midwest ISO is about 8%. The capacity credit from the EWITS study with aggregated wind generation is about 25%. If the wind generation for a 20% US Eastern Interconnection case is 230,000 MW, then 39,000 MW less of some type of generation would be needed to serve the load. If the cost were about \$600,000 per MW, the aggregation of wind with the transmission overlay would be worth about \$24B. This will not pay for all the transmission, but it would help.
6. Load aggregation would reduce the generation requirement of the US interconnection by about 6%. If the load of the US Eastern Interconnection were 600,000 MW, the reduction would be 36,000 MW. For a savings similar to the wind aggregation number above this will not pay for all the transmission, but it would help.
7. Demand resources can bid to a relatively small area since their contribution is limited by their

small shift factors that contribute to correcting a problem. An overlay transmission system would increase the marketing area for a demand resource to a good part of the US Eastern Interconnection. The probability of being accepted as a bid in the energy or ancillary services market is increased with a transmission system.

8. Reliability is difficult to price. An overlay could provide a means of solving many of the frequency regulation and high impact disturbances that occur today or are expected to occur in the future. Performance improvements may be a choice.
9. Nuclear generation may be an option in a low carbon generation scenario. Nuclear plants are large and are difficult to assimilate into a single utility in a short period of time. A large transmission overlay would allow full assimilation of the output of a large nuclear plant at market prices corrected for transmission and losses.
10. The transmission overlay may also make it possible to use the present brown field sights that nuclear sites presently exist to add additional nuclear generation. The present transmission would have to be expanded greatly for this to be done. Short circuit capacity may be a limitation. If the transmission overlay interconnection were several sections of HVDC, the short circuit capacity could be managed by connecting a generator to an HVDC converter that adds little short circuit capacity and designs the AC connections accordingly. The high-capacity contingency rating of the HVDC overlay could accommodate the loss of a nuclear generator and be stable.

The EWITS and JCSP studies only used the production cost savings from the energy market to justify transmission with a benefit greater than the cost for a 20% wind energy requirement. Typical AC transmission expansions justified for reliability criteria have a benefit to cost ratio of about 0.3–0.6 to 1. Grouping lines together that have synergistic affects can raise the benefit to cost ratio near one to slightly greater than one. Designing a system of transmission lines in a time frame that can load the lines produces a better result for studies with high wind energy.

The evaluation period and method may determine the outcome as much as the study. Wind and other

renewable resources have costs that do not escalate significantly with time. The fuel and operation and maintenance costs for conventional generation escalate with time. Thus, a short-term evaluation tends to minimize the value of renewable resources.

In the USA, most wind generation Renewable Portfolio Standards will not reach their maximum percentages until 2025. Multiple HVDC transmission lines are included in the designs to transfer the power associated with the RPS. The Midwest ISO alone has a 23,000 MW RPS level by 2025. The minimum self-contingent design for a three 800 kV HVDC line system is 14,300 MW. A transmission overlay would be possible with the Midwest ISO generation alone. If shorter times are chosen for the study, the generation to justify a transmission overlay is not available.

The seven step transmission planning method used by the Midwest ISO, EWITS, and the JCSP uses a top-down method of planning (see Fig. 15). The end result is calculated first for the entire system. This is often referred to as “holistic” planning. The transmission overlay is planned as a system and not as a series of lines with this method. The generation necessary to load the entire system is evaluated. The sequence of getting to the end state is calculated next. Experience has shown that the larger the area studied the better the results. The largest studied is the Eastern Interconnection with a focus on the US Eastern Interconnection. The end result cannot be constructed as the sum of smaller area solutions in the large area. The Midwest ISO performed Midwest ISO only studies. Very little transmission was required as the entire area has about the same type of choices for generation with just a small change in fuel prices from the east side of the system to the west. Benefits less than the cost of transmission were obtained from the study. If such a transmission expansion were constructed, the higher cost of the Midwest ISO component would just raise the cost of the whole project and reduce the benefit to cost ratio of the transmission overlay.

Bottom-up processes start at the present and add lines sequentially in time. The bottom-up process is an extreme case of sub-optimization. Short times are used that limit the possibility of choice of higher efficiency multiple line systems. More money in total is spent on the sequential expansion than a coordinated expansion. Transmission funds are used to mitigate

constraints with limited value as the constraint just moves to another location unless multiple lines are constructed. The main argument in supporting the bottom-up method is that this is the way it has always been done and will be done as the state regulatory processes also approve one line at a time.

The Rule of Three explains the difference in the results in sequential bottom-up study processes and a top-down system design.

The Rule of Three is a very simple economic concept that applies to most transmission choices.

1. If one line is to be constructed, then the voltage probably should be of the highest existing voltage if the line can be loaded. The reason for this is that the present transmission lines limit the loading on a higher voltage line well below the capability of the line. Bottom-up sequential processes are always in this mode. Some organizations like the status quo and argue for bottom-up sequential processes.
2. If two lines are to be constructed, the difference between a higher voltage overlay and an existing voltage is about breakeven. The second higher voltage overlay increases the load that both higher voltage lines can carry. There seems to be an exception of this with long HVDC lines. Two lines are enough to be more economical. The problems with loop flow spreading benefits over the AC system, even to entities that probably would not care to participate in the transmission expansion. Free ridership is a problem with long distance AC lines. HVDC lines strategically can place the collection of low-cost energy at a high-priced area.
3. If three lines are to be constructed and are able to be loaded, then the higher voltage lines will be a superior economic choice. Wind generation produces conditions that load lines economically if studied near the full development of the wind resource.

In the USA, there is typically a higher rate of return for generation than for transmission. Utilities have focused on generation and usually supplied the minimum transmission necessary based on reliability and need. The state regulatory mind-set reinforced this decision mode. The higher rates of return for FERC-approved (interstate) transmission have changed the decision mind-set of the utilities. Independent

transmission companies are willing to address multi-line transmission expansions if they can obtain tariff cost recovery from an RTO.

Europe with the difference of hydro generation in the Nordic area, coal in Germany, the nuclear in France, the solar potential in Spain, and the distributed wind may also have the potential for a similar high-voltage transmission overlay.

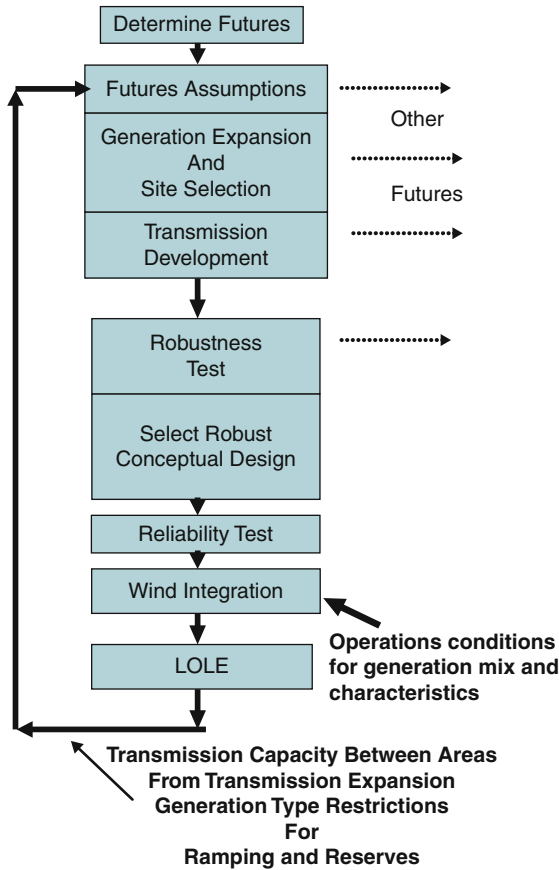
At the power levels being identified and the high cost of overhead transmission it may be justified to the underground HVDC overlay or at least major portions of it. Socially, underground HVDC may be an acceptable solution while overhead may be very difficult to impossible.

Value-Based Long Range Planning Method

The Midwest ISO planning method starts with an economic analysis to develop the conceptual transmission expansion for each future or scenario. An optimum generation expansion plan or forecast is produced for each future. Generation mixes and locations affect the transmission design. A transmission plan is developed for each future based on the economic information from studies. The conceptual plan will be economical. Since there are multiple futures and multiple conceptual transmission plans one looks for transmission lines that are common to every plan or could be modified to fit the other plans. The transmission is the “no regrets” transmission. No matter which future actually happens, the transmission will still be needed. The common plan is tested for robustness with about four other futures to determine the transmission that would be added to the “no regrets” plan to work well with that future.

Figure 15 is a process diagram of the value-based transmission planning method.

The futures are coordinated with the Planning Advisory Committee (PAC). The PAC consists of sector representatives from the transmission owners, independent transmission companies, generators, environmental, large customers, regulators, etc. A future is a set of assumptions that describes a possible outcome. A future has its own generation forecast. A 20% wind energy future may be one future. An all-gas conventional generation may be a second future. Delayed transmission may be a third future. Nuclear forced in



Wind Power Grid Integration: Transmission Planning.
Figure 15

The seven step transmission planning method used by the Midwest ISO, EWITS, and the JCSF

at a given level may be a fourth future. A Reference future usually is the most economic generation forecast without future constraints.

Generation forecasts have an excess of transmission capacity due to the fact that wind generation does not produce energy at the peak periods reliably. Wind generation in the Midwest ISO footprint presently is given an 8% capacity credit to apply to the peak load period. The other 92% is excess generation capacity that may produce energy off-peak.

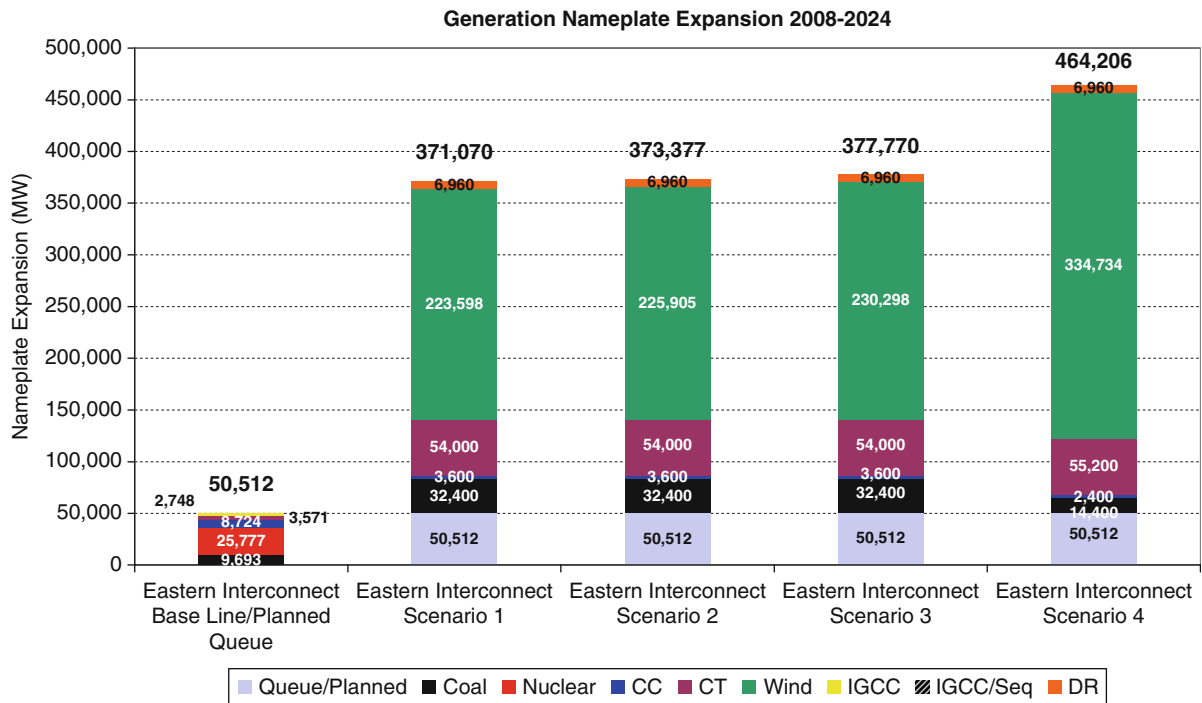
Sensitivity studies are run on futures. A sensitivity study may be a carbon tax. A carbon tax could also be a future with its own generation expansion. Different wind site locations may be sensitivity. Different wind sites may require a unique transmission expansion plan per sensitivity.

Figure 16 shows the generation forecasts for the EWITS futures. Scenario 1 is a 20% wind energy future with generation located strongly in the western part of the Eastern Interconnection. Scenario 3 is a 20% wind energy future with wind generation strongly located offshore of the east coast of the USA. Scenario 2 is a mixture of Scenarios 1, 3. The 20% futures have about 225,000 MW of wind generation. Scenario 4 is a 30% wind energy future, with 330,000 MW of wind generation. The capacity factor of the wind changes the generation that is needed to produce 20% wind energy.

Figure 17 shows the location of the generation by type for Scenario 2 of EWITS. The location of the generation affects the transmission system design.

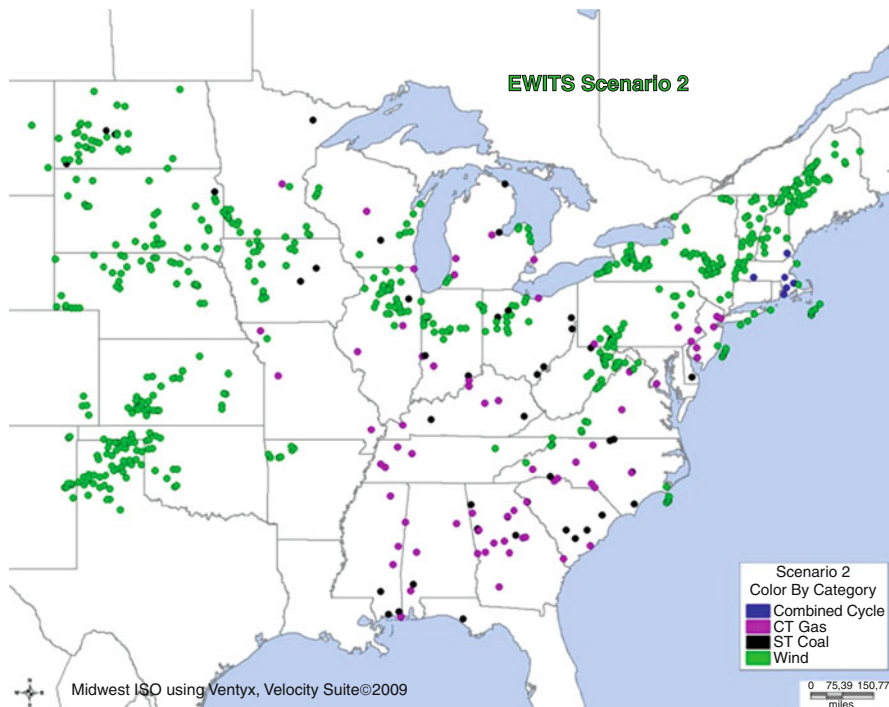
Information to design a transmission expansion is determined from differences in the 8,760 h per simulated year of market operation runs. A simulation includes the economic modeling of every generator in the Eastern Interconnection, the existing transmission, and load data. A Reference future can be run with its own transmission expansion or without the transmission expansion. The with transmission expansion would be used in the “no regrets” robust transmission selection process. The without transmission expansion would be used as a starting point for each of the other scenarios. The without transmission expansion run uses the present transmission with a future generation forecast. The without transmission expansion is referred to as the constrained case. A second case is run that is called a “copper sheet.” A copper sheet has no constraints to power flow. Energy flows to where it would in a perfect energy market. Existing transmission line ratings are placed at very high values. No resistance is modeled for the transmission lines. No interchange flow is modeled. No hurdle rates (the pancake transmission service rates between areas or the inefficiencies in the energy markets) are modeled.

The wholesale energy price differences with the expansion forecast generation without transmission expansion is shown in the left plot of Fig. 18. The right-hand plots are the whole sale with the transmission expansion (Fig. 19). There is just a little congestion remaining with the transmission. Some congestion must remain to produce a value for the transmission.



Wind Power Grid Integration: Transmission Planning. Figure 16

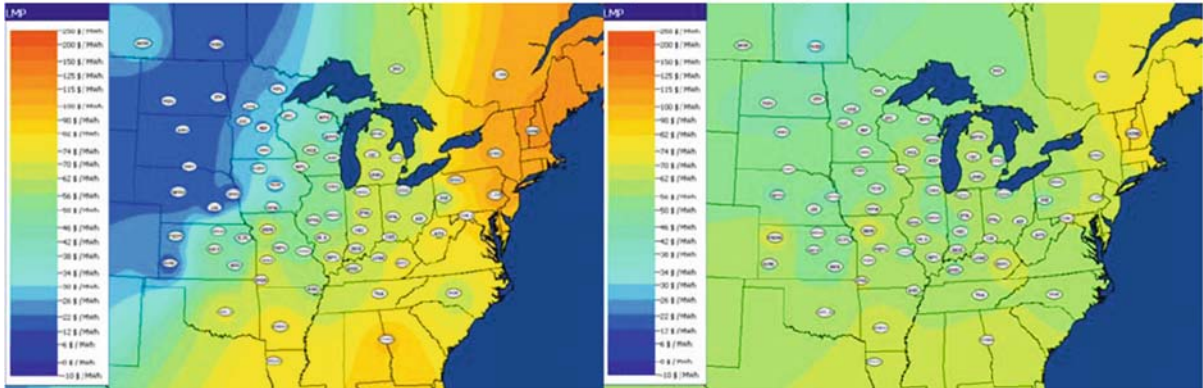
Generation forecasts for the EWITS futures



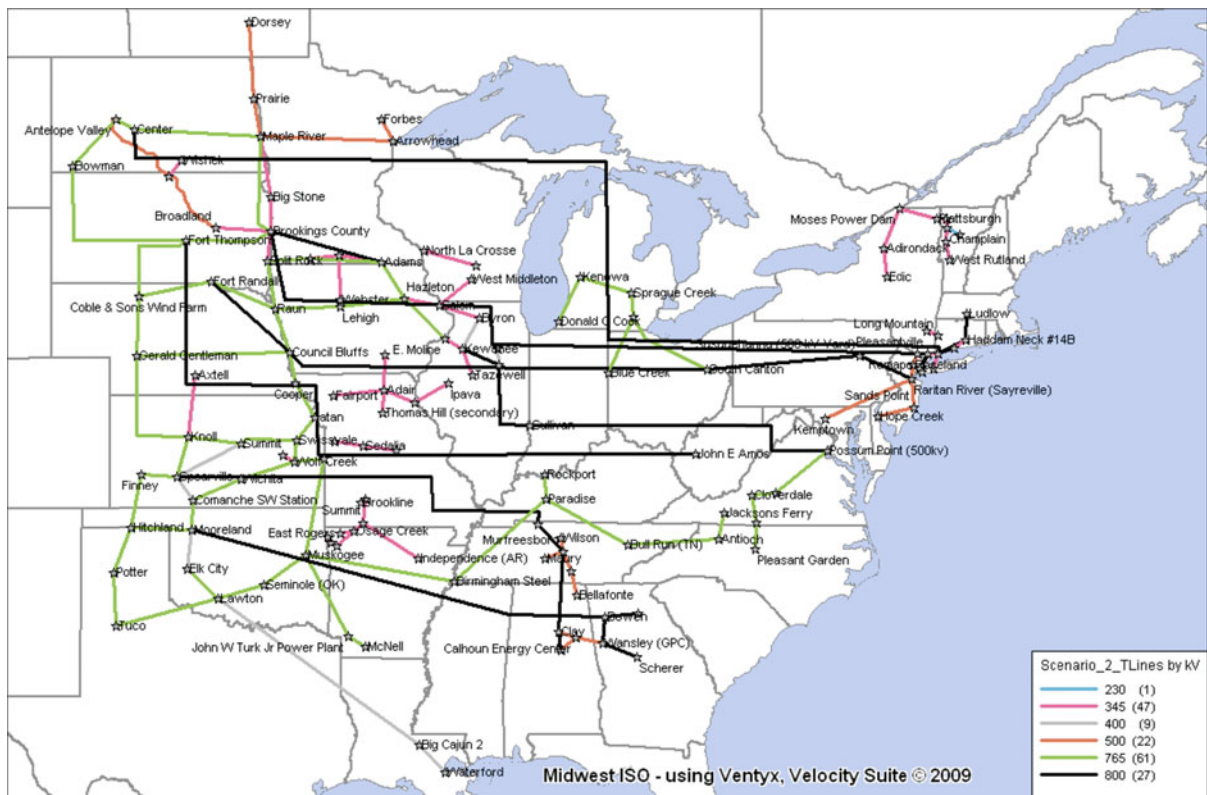
Wind Power Grid Integration: Transmission Planning. Figure 17

The location of the generation by type for Scenario 2 of EWITS

Regional Generation Weighted LMP Changes



Wind Power Grid Integration: Transmission Planning. Figure 18
Scenario 1 annual generation weighted LMP comparison



Wind Power Grid Integration: Transmission Planning. Figure 19
Resulting transmission expansion

The constrained price as shown in Fig. 18 on the left shows the price difference that may exist before the transmission overlay is constructed. Generator pockets (blue) and Load pockets (orange) can be easily found.

The changes in the color represent areas of constraints called congestion. Congestion is a measure of the inefficiency of an energy market. Quebec is not modeled and the chart uses New England's prices. Price is one

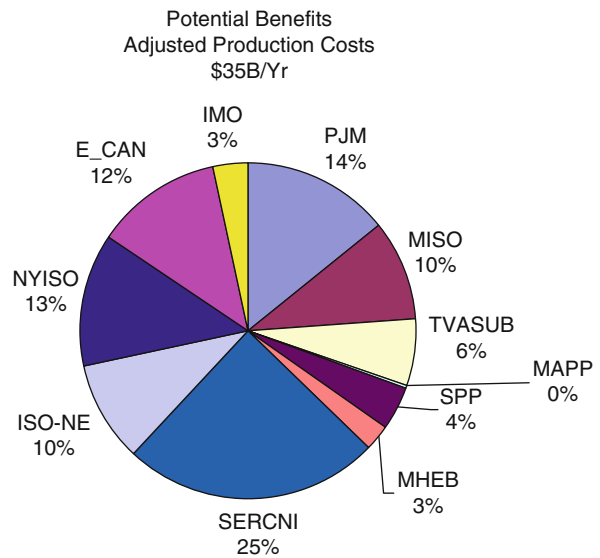
tool to use to determine the transmission, but a better indicator is where energy is sourced and where energy is sunk from the economic models. Sources and sinks include both price and the quantity of energy which are needed to determine value. A price differential for a volume of energy produces revenue that can help pay for the transmission expansion. A high price with a low volume of energy associated with it may have a lower value than a lower price with a high volume. While price is a market signal for the energy markets, it is not a good signal for transmission planning.

The difference in the production cost, the Constrained versus the Copper Sheet, adjusted for sales and purchases provides an estimate of the total potential budget for transmission. Experience has shown that a practical budget may be 70% of the maximum potential budget. Figure 13 is an example of the maximum potential budget for transmission expansion broken down by area. Geographical areas in the USA could be geographical areas in another large country or countries such as in Europe.

The pie chart shows the benefits to the loads. There are also benefits to the generators. Depending on the regulation agreement, sometimes the total benefit to an area is a sum of the load price changes and the generator revenues. If an area is a low-cost producer, the generation can be self-scheduled to serve your own load to prevent increases in the price of energy to the load. The generation schedule does not change the self-scheduling if an area is a low-cost producer is just an accounting process. The results of the study do not change. These are real-time details that cannot be modeled 15 years in advance, but they should be considered when reading a result Figure 20.

Price differences, while indicative, cannot pay for transmission lines. A volume of energy at a good price difference is needed. The location of where a transmission line would collect low-cost energy and where the transmission line would deliver the low-cost energy to a high-priced area can be determined by looking at the increased energy supplied from an area's generation. Figure 21 is a plot of sources (dark red) and sinks (dark blue) that represent the direction of major transmission expansions as indicated by the arrows.

Figure 22 is a bubble diagram or interchange flow diagram that shows the difference in the flow of energy



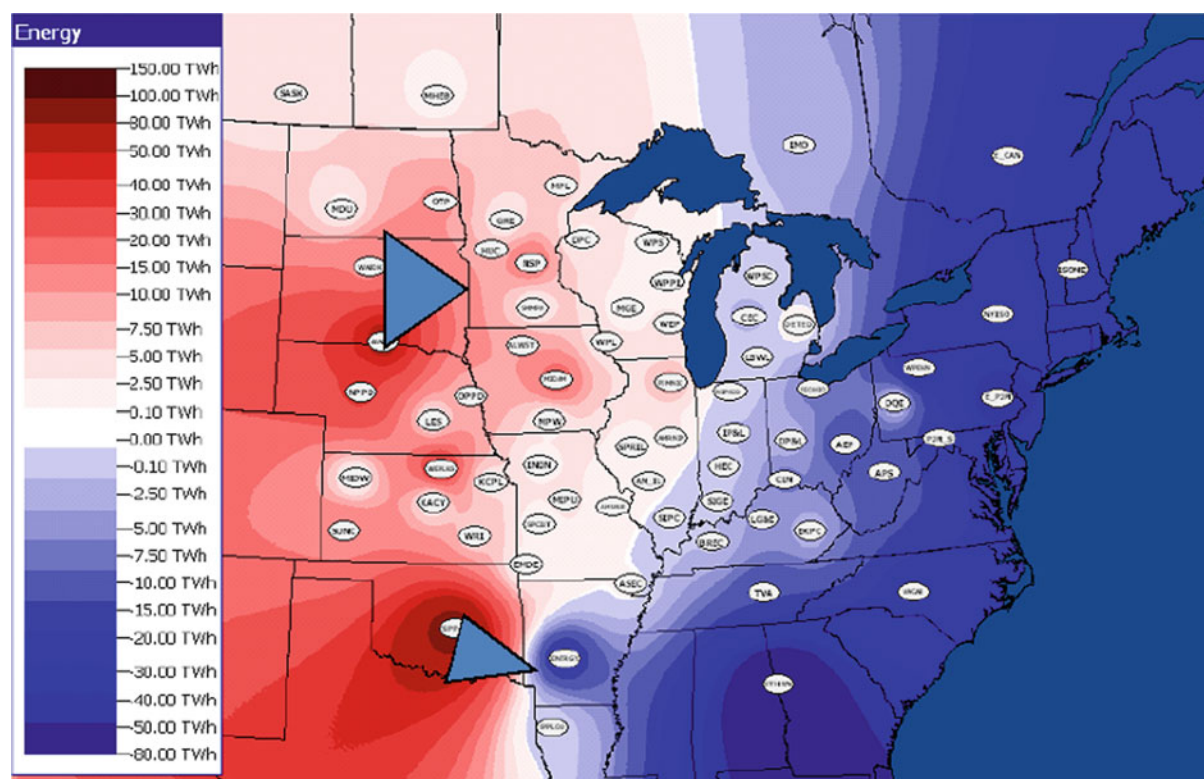
Wind Power Grid Integration: Transmission Planning.

Figure 20

From the EWITS study

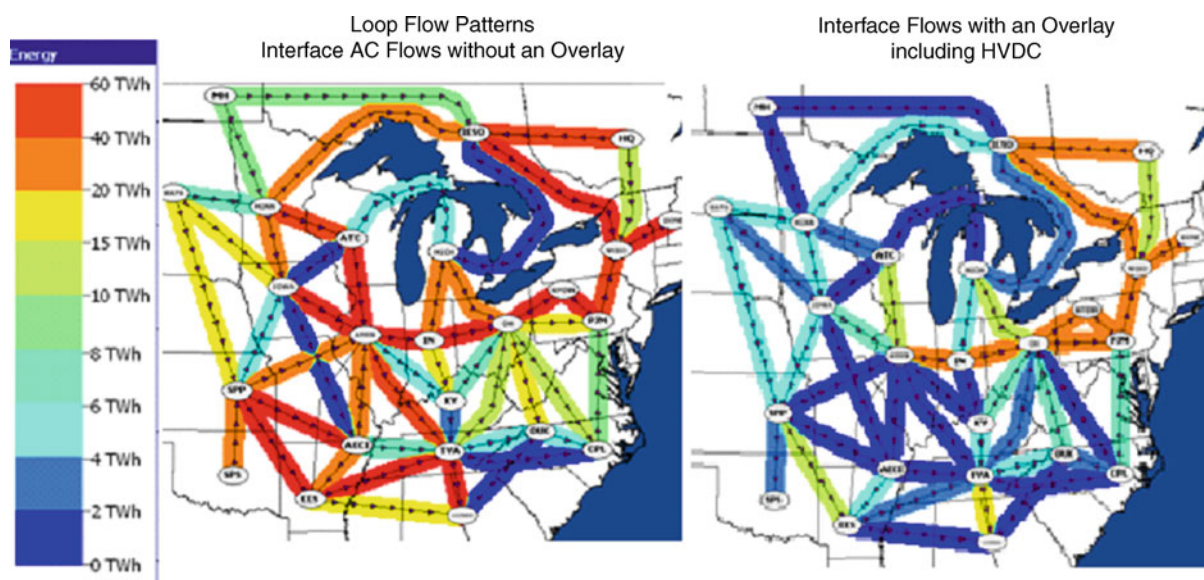
in the Copper Sheet case versus the Constrained case. The general direction of transmission and a relative size can be determined from this diagram. The diagram on the right is a difference with the Copper Sheet and the Overlay. The remaining transmission and direction is indicated by the lines. Most of the incremental power flow has been delivered directly by the Overlay. The areas that may use more transmission are in *orange*. The blue color indicates little need for additional transmission capacity between the areas. The blue also indicates the lack of loop flow through the underlying transmission system. The underlying transmission system would probably require few expansions due to the delivery of the wind energy. The underlying system will not be burned down due to the Overlay as some suggest. The use of a self-contingent HVDC and AC Overlay allows the control of the loop flow by using the scheduling ability of the HVDC lines.

The hourly economic simulations produce the hourly power flow on transmission lines for a year. The black heavy line on the left interchange diagram shows an interface which is the sum of flows on lines crossing a geographic border. Interfaces are defined perpendicular to major flows at areas of change of



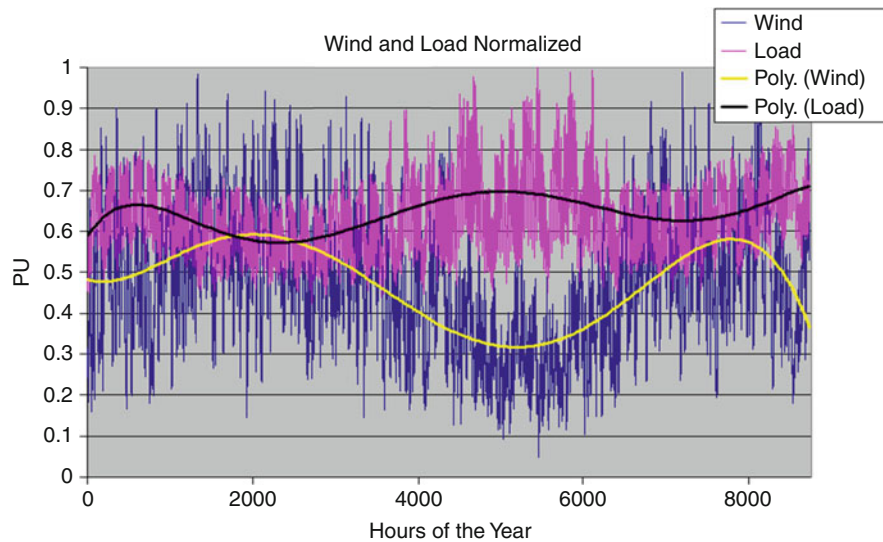
Wind Power Grid Integration: Transmission Planning. Figure 21

Scenario 2 generation difference between unconstrained case and constrain



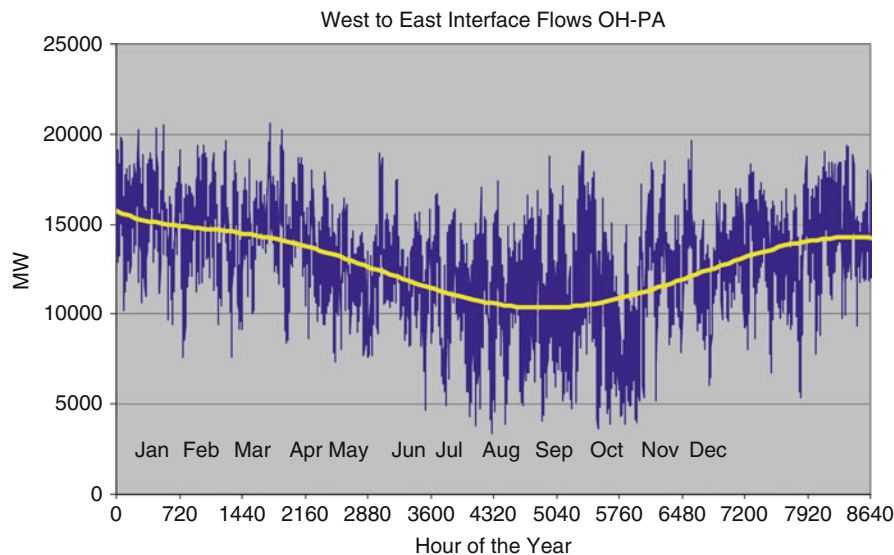
Wind Power Grid Integration: Transmission Planning. Figure 22

Bubble diagram or interchange flow diagram that shows the difference in the flow of energy in the copper sheet case versus the constrained case



Wind Power Grid Integration: Transmission Planning. Figure 23

Wind generation output versus the load



Wind Power Grid Integration: Transmission Planning. Figure 24

Load versus interchange

flow. Locations can be estimated by the places that colors change on the price plot (Fig. 18) or the source sink diagram (Fig. 21). The heaviest loading of the transmission is off-peak in the spring, fall, and winter. The lightest loading of the transmission system is during the summer peak. The plot of wind generation

output versus the load is shown in Fig. 23. The trend lines through the plots show that wind energy production is about to counter when it is needed most. The interchange plot in Fig. 24 shows a similar pattern. The interchange plot shows that the heaviest transmission flows occur off-peak. The spare

transmission capacity on peak adds to the generation reliability. Generation can use the spare transmission on peak to transfer power into an area that would need it. The amount of transmission that is needed can be estimated by inspection, but a more formal method is used.

Loop flow data can be sorted from ascending to descending values and then a load duration curve of the power flow across an interface can be plotted. Based on this curve one can estimate how often a certain transmission level is exceeded, and then a selected level of transmission capacity can be selected. The energy flow at the maximum power level would be very small and create little value. Experience has shown that if a transmission line that would transmit 80% of the energy, not 80% of the peak value, then the line would have good potential for being an economical expansion, c.f. Fig. 25. The load duration and 80% energy process converts energy flow to power transfer (MW) quantities. Transmission lines are rated in MW of power transfer capability. Once the power transfer level is known, the number and voltage of transmission lines can be estimated for each interface.

An examination of the flow that transfers power through multiple interfaces provides an estimate of the

distance that power flows through areas and may be a candidate for a longer AC line or may qualify for a DC line if the distance is more than 600 miles in the examples given previously. Table 1 shows the power transfers through multiple interfaces for the EWITS futures.

Power transfers between multiple interface locations can be used to estimate the transmission that would be included in the overlay.

Table 1 shows the interface numbers using this process for the EWITS study. Scenarios 1–3 had a 20% wind energy level with about 225,000 MW of wind generation. Scenario 2 had a mixture of wind generation using onshore and offshore wind. Scenario 1 used primarily high-capacity factor wind located in the high plains in the western part of the Eastern Interconnection. Scenario 3 had a high proportion of the wind offshore of the east coast. Scenario 4 had a 30% wind generation scenario with strong wind both offshore and onshore in the west as well as in the middle for about 330,000 MW of wind generation. Energy is converted to power transfer capability in MW which is what transmission lines are rated. The Ohio-Eastern PJM (OH-EPJM) Scenario 2 or compared to a three-line HVDC system rating of 14,300 MW one can see it is about a match. The distance across the system is greater than 600 miles. So the HVDC lines terminals are selected using the information in the source-sink diagram in Fig. 21. The choice of the least expensive option in Fig. 12 further validates the choice. Another simulation is chosen to simulate your choices. The examination of the constraints and the information

← Transmission Capacity designed to deliver 80% of desired energy flow

Wind Power Grid Integration: Transmission Planning.
Figure 25
The 80% energy process

Wind Power Grid Integration: Transmission Planning. Table 1 Top 24 interfaces with the largest annual energy difference

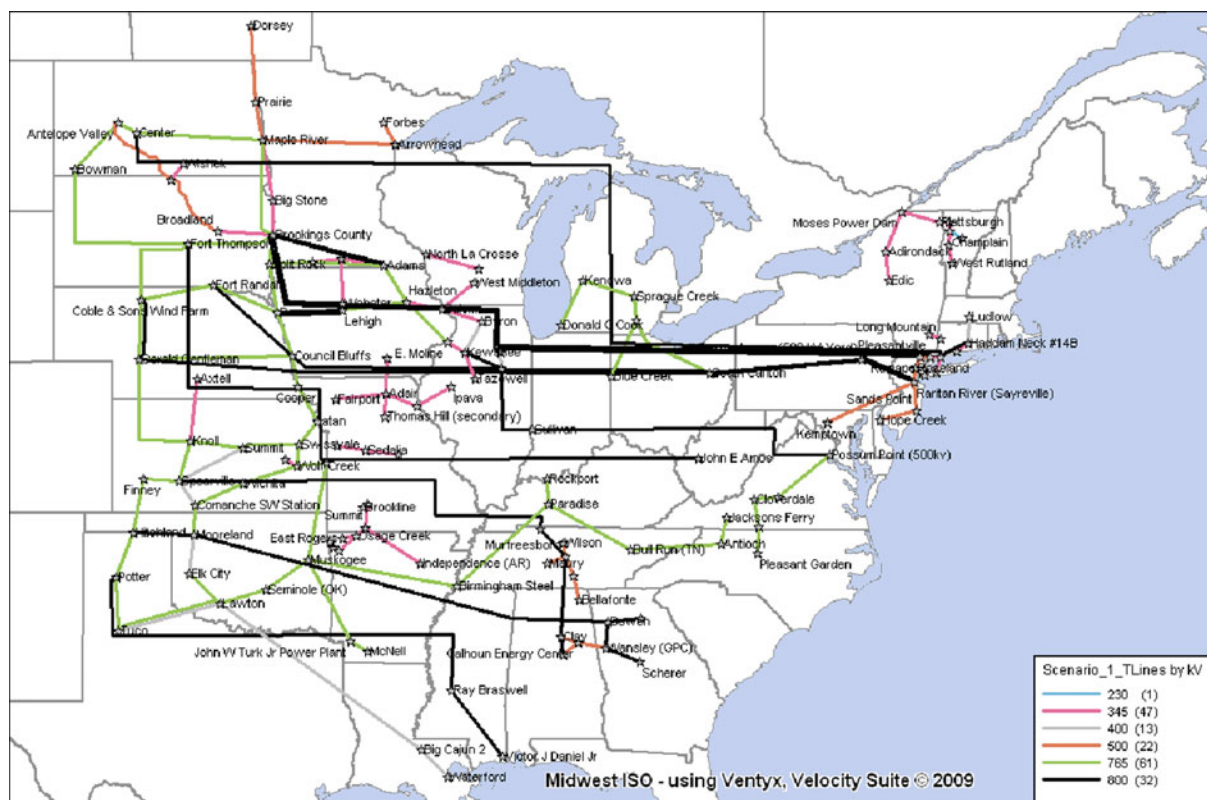
Interface	EWITS Scenario 2	EWITS Scenario 1	EWITS Scenario 3	EWITS Scenario 4
	Additional transfer needs (MW)	Additional transfer needs (MW)	Additional transfer needs (MW)	Additional transfer needs (MW)
AMRN – IN	22,901	28856	13223	24,609
IN – OH	16,594	20843	9633	17,081
OH – EPJM	14,378	18662	8715	12,928
SPS – SPP	13,983	13482	7743	13,399
SPP – EES	11,598	12551	6417	12,160

specific to the new case with the transmission expansion allows the design to be refined to that in Fig. 19. It takes four to five iterations to refine a design. Computer simulations require 6 days to run. Information to design the system is derived from the economic computer simulations and the differences between simulations. It would be very difficult to formulate a set of alternatives using power flow to design a system such as that in Fig. 19.

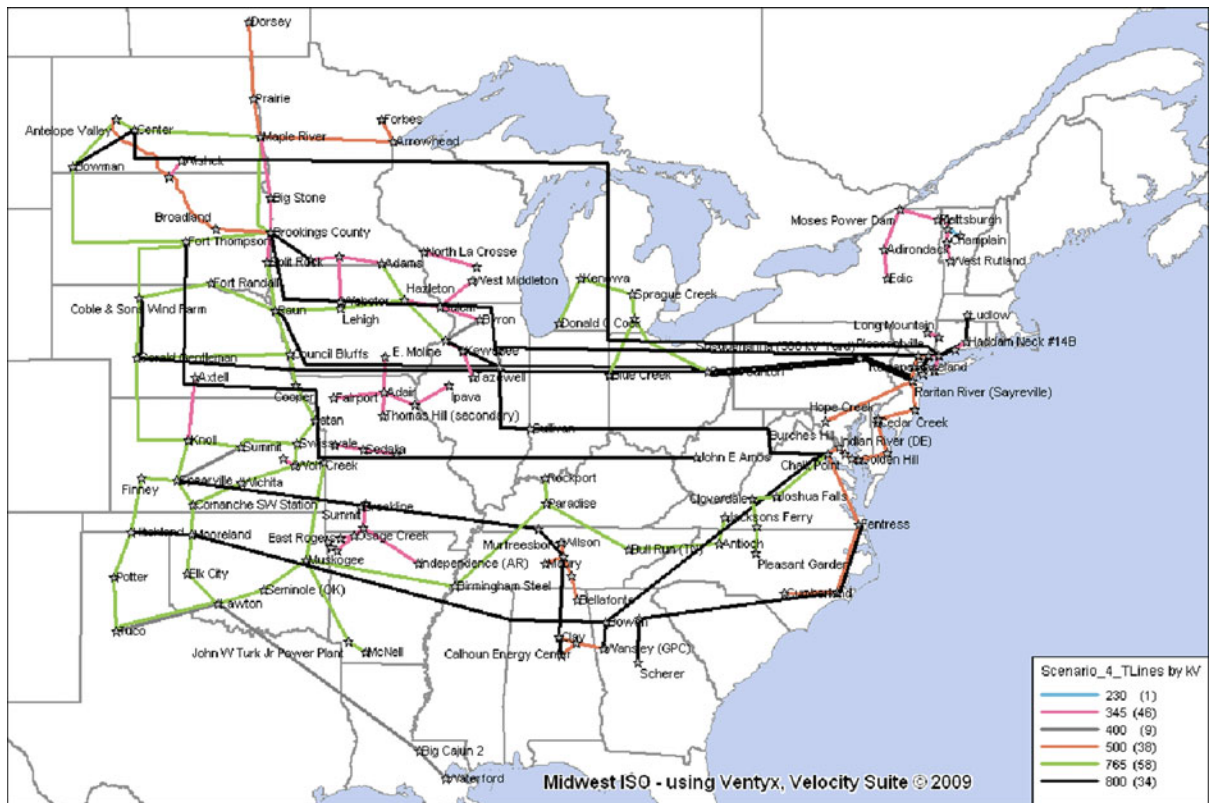
A short visual examination can identify the potential common elements in the four scenarios. A rigorous ranking produces a lot more computer work and does little better to identify a robust transmission plan of “least regrets.” The black lines are the HVDC lines. Green lines are 765 kV lines. Red lines are 345 kV. Orange lines are 500 kV (Figs. 26–29).

EWITS skipped the Robustness Test, The Robust “least regrets” transmission selection, and the

Reliability test. The Reliability section of the study is passing through the least regrets plan developed using economic techniques with the best estimate of the generation, usually a conservative forecast, to determine if the plan could pass the tests that the Midwest ISO subjects all the reliability-based plans. The Midwest ISO added a lot of dynamic reactive devices to the EWITS Scenario 2 transmission plan and loaded the HVDC to maximum generation with wind generation schedules to obtain a solved power flow. General Electric set up a rough dynamics case and ran a few tests of AC and HVDC outages without showing any great concern. Although this process was very rough and would have to be run with many more contingencies, the study gives one faith that it may work with a large wind resource. The Midwest ISO is not about to finance a study to do a definitive dynamics study without the assistance of others for the Eastern Interconnection.



Wind Power Grid Integration: Transmission Planning. Figure 26
EWITS Scenario 1–20% strong western wind generation locations



Wind Power Grid Integration: Transmission Planning. Figure 27

EWITS Scenario 2–20% mixed wind generation locations

The Eastern Interconnection Planning Collaborative will have to do one if it is to be done. If it were done, the wind generation levels would be far less than a 20% energy level.

Figure 30 shows the total cost of the four EWITS futures. The wind capital cost contributes to most of the variation between the futures. Wind integration costs that people obsess about are a small portion of the total cost, even less than the projected variable operation and maintenance cost on the wind machines. Transmission is a distant third in the ranking of components behind Wind Capital Cost and Production Cost (Fuel & VOM).

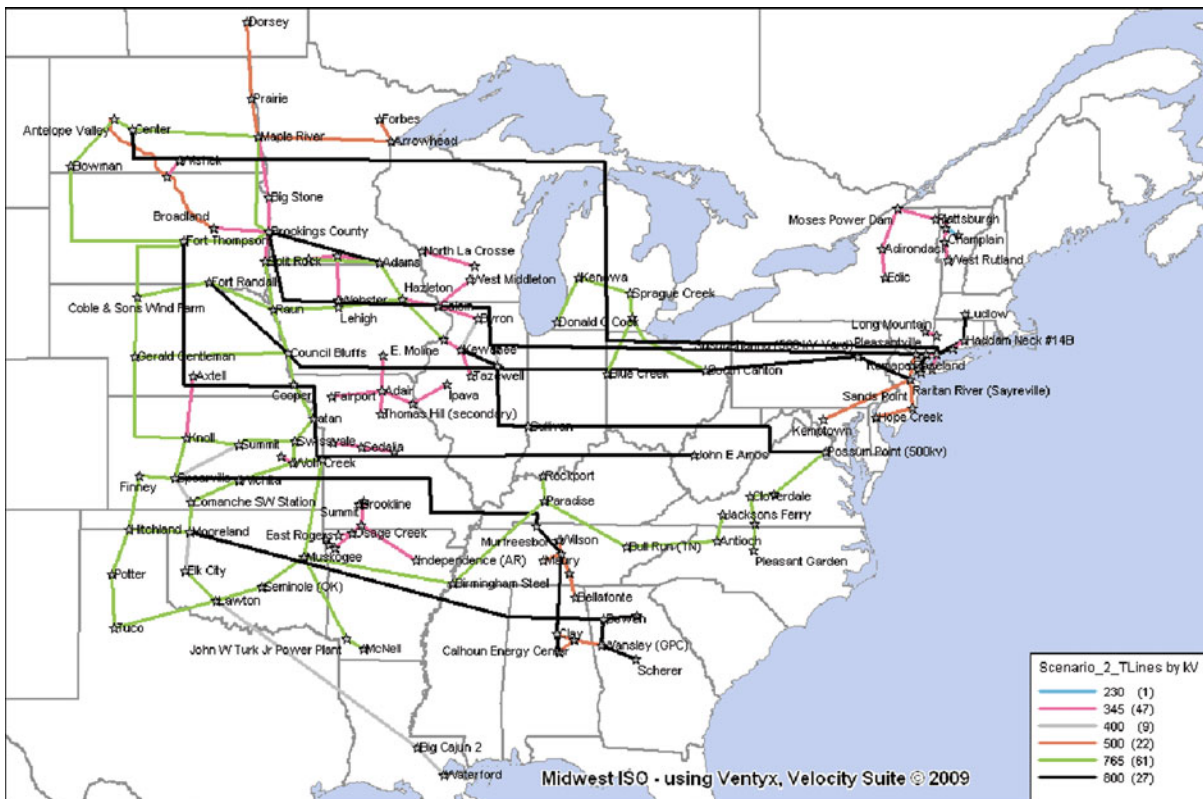
Wind Integration Studies

Statistical distributions of the rate of change of power or ramping requirements were produced. Tails of

the distributions were examined for the maximum rates of change. Generator spinning and fast start reserves were estimated to manage the tail events. The reserves in the economic simulations were changed in the dispatch algorithm to model the system operation with the higher reserves. The cost of operation with a perfect wind forecast and a forecast with forecasting errors were compared to determine the cost of wind integration.

The output of the wind integration section of the study can be used to modify the assumptions to be used in the next iteration of the study. Spinning reserve, fast start requirements, and the capacity credit for wind are three key factors that modify the generation forecast assumptions list and available generation lists.

Probabilistic or statistical planning techniques with multiple year runs are needed to assess the cost of variability and the effects on transmission system



Wind Power Grid Integration: Transmission Planning. Figure 28

EWITS Scenario 3–20% heavy offshore east coast wind generation locations

design. Should one design a transmission system for the best wind year? It may be lower in total cost to curtail the extreme wind and build transmission more toward a value that would handle less wind. The economic studies allow one to determine the choice of the level to design transmission for wind energy variability.

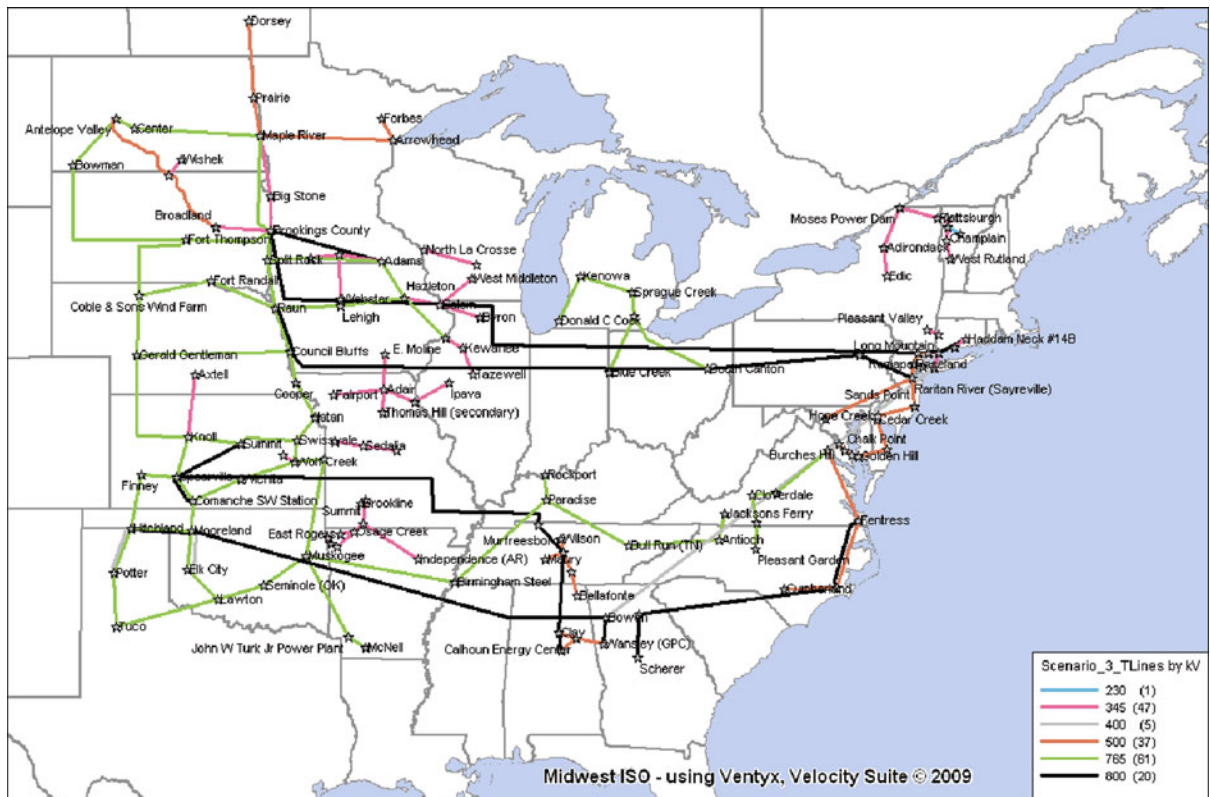
If one must meet 20% energy each year, for example, then there has to be more wind generation added and the associated transmission to deliver a good portion of the wind energy to account for low wind energy years.

The wind energy variations are quite high when taken at the turbine or wind plant level. Geographic diversity of wind resources can be summed by a transmission overlay to produce a less variable wind energy resource than any of the individual plants. The EWITS study produced the graph in Fig. 5 that shows the reduction in variation of wind as the wind is

aggregated. The variation of wind is reduced by a factor of 4 just by aggregation.

One must remember that the weather affects load levels just as it does wind levels. Wind and load patterns must be matched using the same year for studies. Future years must be matched also. Multiple years of wind and load patterns are scaled to future load levels to produce hourly models for economic studies. If statistical models of wind were used, both the wind model and the load model would need to be matched.

Less generation is needed in reserve to manage the wind variation if there is a transmission system that can aggregate wind generation. The transmission lines that aggregate wind energy would have much higher variable loading that would be the case with conventional generation. The reactive power variations would produce voltage variations on the AC power system if the wind were aggregated on the AC system. Dynamic



Wind Power Grid Integration: Transmission Planning. Figure 29

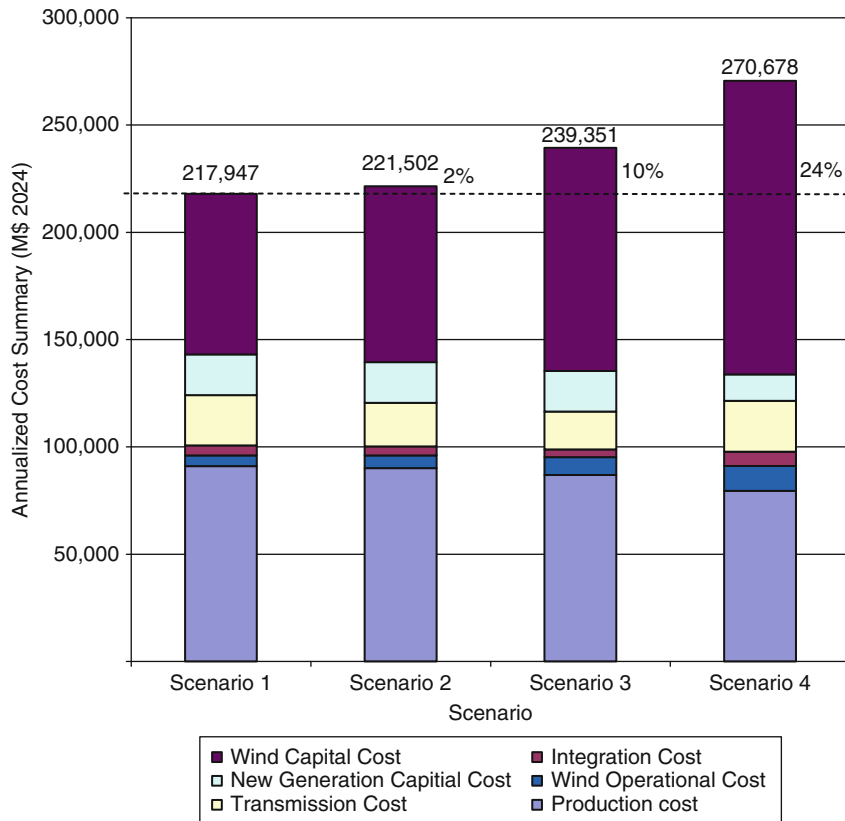
EWITS Scenario 4–30% wind energy, heavy everywhere wind generation locations

reactive power support would have to be distributed across an AC transmission system to manage the voltage in an all AC design. An HVDC transmission overlay would restrict voltage regulation to be at the terminals of the HVDC converters. This would simplify the AC system design. If the HVDC system is scheduled to transmit variability, the areas receiving the variability in energy flow would have to provide dynamic reactive power support. Having the HVDC lines have varying schedules could relieve the AC system from having to have reserve generation close to the wind generation to control frequency and power interchange schedules. Probably gas fired generation would have reduced loading if the HVDC lines were delivering energy to an area. The gas generation on the receiving end could provide the regulation. Ancillary Services revenues would be increased for the generators on the receiving end. Less total generation would have to be constructed to serve wind. More

wind generation could be placed in areas with an HVDC converter than could be accomplished without the HVDC converter. The wind generation system would require fewer AC transmission ties to manage the wind variability.

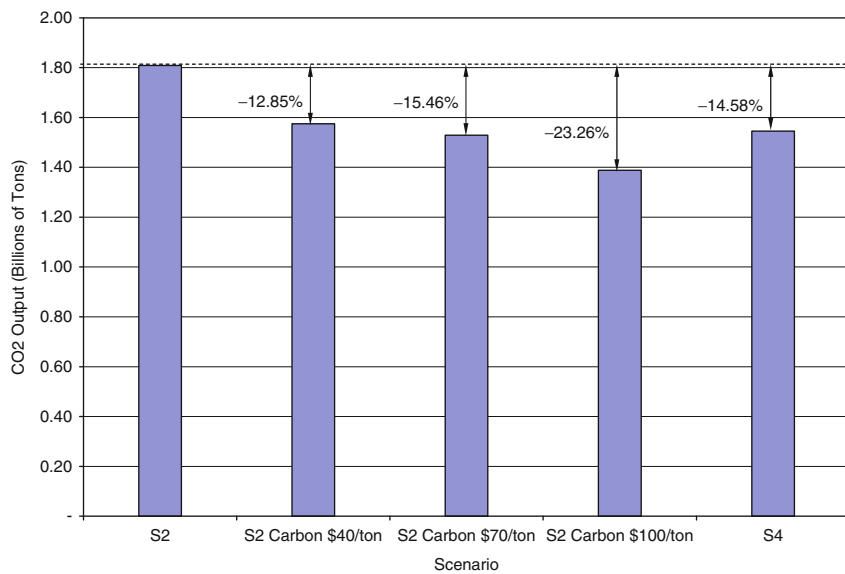
Other Information to Answer Questions

Figure 31 provides information on the carbon dioxide reduction Scenario 2 sensitivity test for various carbon tax prices. The 30% wind energy level reduces carbon dioxide about the same as a \$70/ton tax and costs less. This type of analysis is performed to answer questions about the carbon performance. Other information on the curtailment of wind, conventional generation capacity factors and the time running a minimum generation, the number of starts on generation, the revenue changes on generation, all emissions, etc., can be produced from the economic simulations.



Wind Power Grid Integration: Transmission Planning. Figure 30

Total cost of the four EWITS futures



Wind Power Grid Integration: Transmission Planning. Figure 31

Carbon dioxide reduction Scenario 2: sensitivity test for various carbon tax prices

Conclusion

Overlay transmission systems can be designed to economically and reliably serve load and deliver renewable energy.

Overlays are a layer of transmission which spans the underlying transmission systems. Overlays require different planning techniques than those of the underlying systems.

The criteria and degree of study of the underlying and the overlay vary considerably. The underlying studies are more detailed concerning the design of the transmission facilities. Level of involvement in a number of departments of specialists may occur at a large utility. For a small utility one or a few people may perform many tasks to get the details necessary to construct a transmission expansion.

The involvement of many different people with widely varying backgrounds and interests have driven the transmission planning processes to provide more information than has been required in the past. Larger transmission expansions have attracted more stakeholders in the process.

The creation of the RTO and the insertion of large amounts of wind energy and solar in some areas has changed the way transmission is being planned. Energy resources, such as wind, require probabilistic evaluations that exceed the conventional generation need for such analysis.

Bibliography

1. EWITS – Eastern Wind Integration and Transmission Study. The National Renewable Energy Laboratory, US. http://www.nrel.gov/wind/systemsintegration/pdfs/2010/ewits_executive_summary.pdf
2. EIPC – Eastern Interconnection Planning Collaborative. <http://www.eipconline.com/>
3. JCSP – Joint Coordinated System Plan (JCSP) 2018 summer reliability study Report. https://www.midwestiso.org/Library/Repository/Study/JCSP/JCSP_Report_Volume_2.pdf
4. PROMOD. <http://www.ventyx.com/analytics/promod.asp>
5. GE MAPS – GE Multi-Area Production Simulation Software. <http://crai.com/consultingexpertise/content.aspx?tID=828&subTID=842&terTID=896>
6. GRIDVIEW. <http://www.ferc.gov/eventcalendar/Files/20100608140700-Zhu,%20ABB%20-%206-9-10.pdf>

Wind Power, Aerodynamics and Blade Technology

MARTIN O. L. HANSEN

DTU Mechanical Engineering, Technical University of Denmark, Lyngby, Denmark

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Where Does the Energy Come From

How Is the Energy Harvested

How Much Energy Can Be Harvested

Environmental Conditions

Blade Element Momentum Theory, BEM

Computational Fluid Mechanics

Rotor Blades

Future Directions

Bibliography

Glossary

A Rotor area

a Axial induction factor

a' Tangential induction factor

B Number of blades

C_d Drag coefficient

C_l Lift coefficient

C_p Power coefficient

C_T Thrust coefficient

C_θ Tangential velocity in far wake

D Aerodynamic drag

L Aerodynamic lift

M_R Rotor shaft torque

M Bending moment, torque

P Power

p Load

R Blade radius

r Local radius on blade

T Thrust

u Axial velocity in the rotor plane

u₁ Axial velocity in far wake

V₀ Wind speed

- V_{rel} Relative velocity to blade
 W Induced velocity
 α Angle of attack
 θ Local blade angle with rotor plane
 ρ Air density
 σ Solidity, material stress
 Φ Flow angle
 ω Angular velocity

Definition of the Subject and Its Importance

Wind energy has been used for thousands of years to drive sailboats, for grinding grains, and for pumping water. For approximately 100 years, wind energy has also been used to produce electricity by connecting the rotor shaft to a generator and thus transforming wind mills into wind turbines. With the growing awareness to find alternatives to fossil fuels in order to reduce the CO₂ emission and to become less dependent on oil-producing regions, wind energy has reached a new age, where wind turbines are no longer produced by local black smiths but by very large companies using advanced design tools similar to the aeronautical industry. It is the purpose of this entry to give an introduction to wind turbine aerodynamics and describe the aerodynamic design tools presently used and what tools may be used in the future.

Introduction

Wind turbines are becoming a bigger and bigger contribution to the production of electricity, with global annual growth rates of more than 20%. In windy regions as, e.g., Denmark, 20% of the total electrical energy produced already stems from wind turbines, and this number is even higher in the western part, where the wind resources are larger. In the 1980s, where the commercial wind turbines were introduced to the market, they were considered a nightmare for the utility companies since the electricity production from wind turbines is strongly correlated with the wind speed and thus variable in nature. The utility companies preferred a few large power plants (often coal, gas, or nuclear) that could easily be controlled to balance the consumption. Today they have realized that in windy areas, the price of electricity measured in \$/kWh is actually competitive with the more conventional means of power

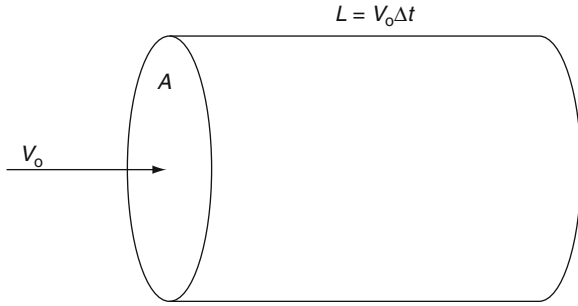
production, and some of them have actually become developers that build and operate large wind farms. The production price from wind turbines has not always been as low as today, and the technological developments done over the 30 years have helped. In the beginning, the fatigue loads from the turbulent wind approaching the rotor were not very well known, and therefore, the structures were built excessively strong using a lot of unnecessary material. Research on wind turbine aerodynamics has helped a lot to reduce the uncertainties of the loads, and thus, it has become possible to reduce the weight measured in kg/kW and thus also the price of the wind turbines. Most of the theories are quite old and stem from propeller theory (ship and airplanes) [1], but have in recent years been extended with empirical models specifically dedicated to wind turbines. A very important contribution to this were two EU projects [2, 3], where most of the available design tools were tested against real measurements and the so-called engineering models calibrated against these full scale experiments. The first part of entry describes very basically the origin of the energy contained in the wind and next some very classical theory is presented as, e.g., described in [1, 4] defining many of the important parameters and non-dimensional numbers used in wind turbine aerodynamics. Next, the classical Blade Element Momentum method (BEM), as derived in [1] but with the necessary engineering models derived in, e.g., [2, 3] is shown. The BEM model is still by far the most used tool in wind turbine design and load calculations, since it is very fast and in most cases quite reliable. Finally, some more advanced numerical tools based on Computational Fluid Dynamics will be addressed.

Where Does the Energy Come From

The energy extracted from a wind turbine comes from extracting some of the kinetic energy contained in the wind. In Fig. 1, A denotes an area perpendicular to the wind velocity.

The mass flow, $\dot{m}$ [kg/s], that travels through this area with a speed of V_o is

$$\Delta m = \rho AL = \rho AV_o \Delta t \Rightarrow \dot{m} = \frac{\Delta m}{\Delta t} = \rho AV_o \quad (1)$$



Wind Power, Aerodynamics and Blade Technology.

Figure 1

Volume of air that has traveled through the area, A , with a speed of V_o in time Δt

where ρ denotes the air density slightly depending on temperature and atmospheric pressure but approximately 1.2 kg/m^3 . For a wind turbine rotor with a diameter of 90 m and wind speed of 10 m/s, the mass flow is around 76 t/s. The kinetic energy, ΔE_{kin} , contained in the air volume shown in Fig. 1 is

$$\Delta E_{\text{kin}} = \frac{1}{2} \Delta m V_o^2 = \frac{1}{2} \rho A V_o^3 \Delta t \quad (2)$$

and thus the available power (provided that the wind speed could be reduced to 0 m/s)

$$P_{\text{avail}} = \frac{\Delta E_{\text{kin}}}{\Delta t} = \frac{1}{2} \rho A V_o^3 \quad (3)$$

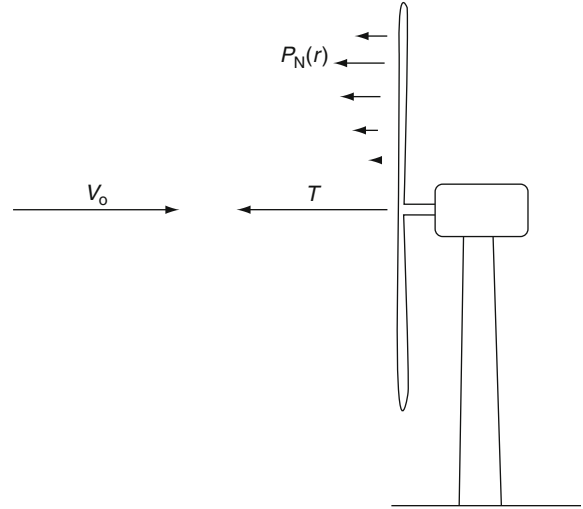
In order to extract kinetic energy from the wind, it is necessary to create a force opposite to the wind velocity and thus reducing the wind speed to something less behind the rotor. This force is denoted the thrust, T [N], and is a result of the loading normal to the rotor disc, $p_N(r)$ [N/m], $T = \int_0^R p_N(r) dr$, see Fig. 2.

How Is the Energy Harvested

The thrust force is created by the local flow past the rotor, which resembles a man cycling with the wind coming 90° to his heading, see Fig. 3.

In Fig. 4 is shown a similar situation, where the velocity of the bike is replaced by the rotational speed of the rotor, $\omega \cdot r$.

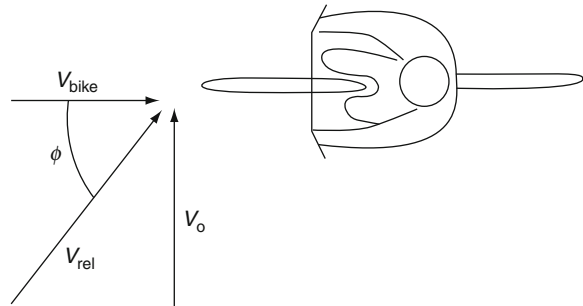
Comparing Figs. 3 and 4, it is seen that the only difference is an extra velocity, $\mathbf{W}$, appearing on the blade section. This velocity is denoted the induced velocity and is caused by the aerodynamic forces



Wind Power, Aerodynamics and Blade Technology.

Figure 2

Thrust force stemming from normal load distribution to slow down the wind speed

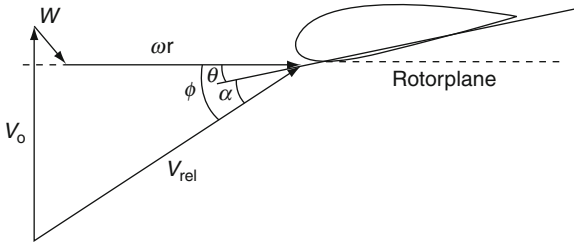


Wind Power, Aerodynamics and Blade Technology.

Figure 3

Man cycling with the wind speed coming 90° to his heading

created by the flow past the blades. The relative wind, $\mathbf{V}_{\text{rel}}$, seen by a blade section is thus the vector sum of the wind speed, $\mathbf{V}_o$, the induced wind, $\mathbf{W}$, and the rotational velocity due to blade rotation, $\omega \cdot r$. If the blade section shown in Fig. 4 is mounted at an angle to the rotor plane, θ , then the flow past the blade section is similar to the flow past an airfoil with an angle of attack of α as shown in Fig. 5. Φ is denoted the flow angle, and the angle of attack is the difference between the flow angle, Φ , and θ .



Wind Power, Aerodynamics and Blade Technology.

Figure 4

Flow seen on a local section of the blade. The situation resembles the cycling man seen in Fig. 3, except for the induced velocity W

From Fig. 5, it is directly seen that the normal load stems from the lift and drag projected normal to the rotor plane as

$$p_N = L \cos \phi + D \sin \phi \quad (4)$$

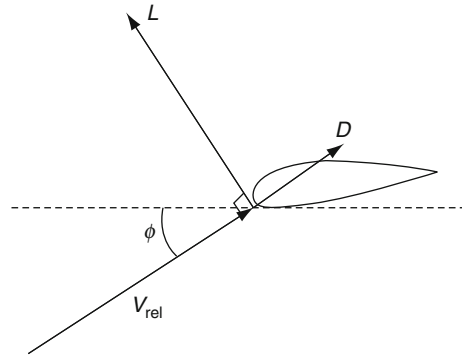
Also there is a tangential load, which can be found as the lift and drag projected into the rotor plane as

$$p_T = L \sin \phi - D \cos \phi \quad (5)$$

The induced velocity shown in Fig. 4 is a direct result of the aerodynamic load that also influences the incoming air. To conclude, the thrust needed for extracting kinetic energy from the air is created by the normal component of the aerodynamic loads and the tangential loads yield a torque on the rotor shaft, M_R , that can be used to drive a generator as

$$M_R = \int r p_T dr \quad (6)$$

From Fig. 4, it is seen that the angle of attack, α , is increased with increasing wind speed provided that the angular velocity is kept constant. The lift force is for small angles of attack increasing almost linearly with the attack α , since the airflow following the aerodynamic profile is more and more bent requiring larger pressure forces. For these low angles of attack, the drag force is small and relatively constant. There is, however, a limit on how much the flow can be bent, and at a certain angle of attack, the flow will no longer be able to follow the surface of the airfoil but will separate and large vortices are formed on the downwind side of the wind turbine blade. In this situation, the flow is said to stall and the lift is decreased and the drag is quickly



Wind Power, Aerodynamics and Blade Technology.

Figure 5

Aerodynamic forces on an airfoil with an angle of attack of α with the incoming wind

increased, reducing the tangential load, see Eq. 5. When the tangential load is decreased also the rotor torque is reduced, according to Eq. 6, and thus also the power. This is exploited on so-called stall-regulated wind turbines where the blades are rigidly clamped to the hub and the angle with the rotor plane is thus constant. If the angular velocity at the same time is constant, then the wind turbine will stall at a certain wind speed and thus passively reducing the power at high wind speeds.

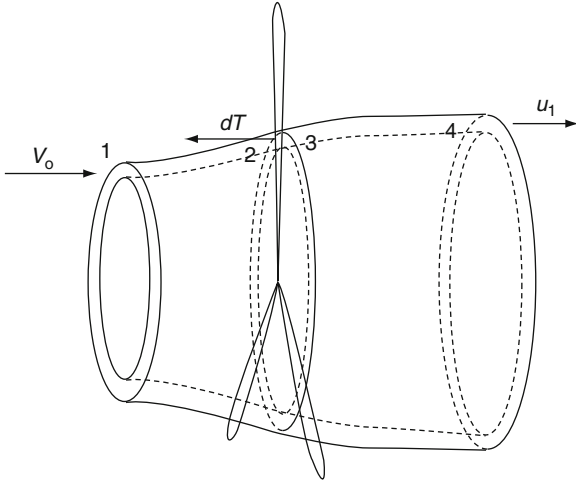
How Much Energy Can Be Harvested

To answer this question, one must consider an ideal rotor without losses. Figure 6 shows a streamtube intersecting the rotor plane at radius, r .

The conservation of linear momentum on the control volume comprised by the streamtube and neglecting the axial component of the pressure forces yields the local contribution, dT , to the total thrust

$$dT = \rho \cdot u \cdot dA (V_o - u_1) = 2\pi r \rho u (V_o - u_1) dr \quad (7)$$

where $\rho u dA = 2\pi r \rho u dr = d\dot{m}$ denotes the mass flow through the rotor. dA is area of the strip in the rotor plane, u the velocity in the rotor plane, and u_1 the velocity in the far wake. Further, the Bernoulli equation can be used from point 1 to 2 and again from point 3 to 4 in Fig. 6, yielding that the pressure jump across the rotor can be expressed as



Wind Power, Aerodynamics and Blade Technology.
Figure 6

Streamtube of the flow through the rotor

$$\Delta p = \frac{1}{2} \rho (V_o^2 - u_1^2) \quad (8)$$

Utilizing that the thrust force stems from the pressure jump across the rotor as $dT = dA \cdot \Delta p$ Eqs. 7 and 8 give that the velocity in the rotor plane is the mean of the undisturbed wind speed and the speed in the far wake

$$u = \frac{1}{2} (V_o + u_1) \quad (9)$$

For more details about the derivations of Eqs. 7–9, see [5]. Due to the tangential loads, Eq. 5, the wake is set into rotation and the relationship between local torque, dM , and tangential velocity in the far wake, C_θ , can be estimated using the conservation of angular momentum and since the power is the torque multiplied with angular velocity, the local contribution to the power becomes

$$dP = \omega dM = \omega \cdot d\dot{m} \cdot r \cdot C_\theta = 2\pi r^2 \omega \rho u C_\theta dr \quad (10)$$

The axial, a , and tangential, a' , interference factors are introduced as the non-dimensional axial and tangential induced velocities in the wake just behind the rotor as

$$a = \frac{W_{\text{axial}}}{V_o}$$

$$a' = \frac{W_{\text{tangential}}}{\omega r}$$

The induced velocity, $\mathbf{W} = (W_{\text{axial}}, W_{\text{tangential}})$, is drawn in Fig. 4. Equations 7 and 10 are derived as if the loads, tangential and normal, are distributed evenly on the annulus at the rotor plane sketched in Fig. 6. This situation corresponds to a wind turbine with an infinite number of blades, and since this is in practice, it is not the case Prandtl derived a correction to the momentum equations by introducing the so-called Prandtl tip loss factor

$$F = \frac{2}{\pi} \cos^{-1} \left(\exp \left[-\frac{B}{2} \frac{R-r}{r \sin \phi} \right] \right) \quad (11)$$

where B is the number of blades, R the rotor radius, r the local radius, and ϕ the flow angle at radial position r . The correction is applied by multiplying the right hand sides of Eqs. 7 and 10 by F . More information on the derivation of Prandtl's tip loss correction can be found in [1].

Utilizing this correction and applying expression (9) for the axial velocity at the rotor plane together with the definitions of the axial and tangential induction factors yields an expression for the available power. Note the induced velocity in the far wake is twice that of the one in the rotor plane, i.e., $C_\theta = 2a'\omega r$

$$dP = 4\pi \rho \omega^2 V_o a' F (1-a) r^3 dr \quad (12)$$

Equation 12 can be integrated to

$$P = 4\pi \rho \omega^2 V_o \int_0^R a' F (1-a) r^3 dr \quad (13)$$

Now introducing the power coefficient, C_p , as the ratio between the shaft power and the maximum available power

$$C_p = \frac{P}{\frac{1}{2} \rho V_o^3 A} \quad (14)$$

and the tip speed ratio, λ , as the ratio between the rotational speed of the tip and the free wind speed

$$\lambda = \frac{\omega R}{V_o} \quad (15)$$

Eq. 13 can be expressed in terms of the power coefficient and the tip speed ratio as

$$C_p = \frac{8}{\lambda^2} \int_0^\lambda a' F (1-a) x^3 dx. \quad (16)$$

So to optimize the power coefficient, the expression $f(x) = a'F(1-a)x^3$ should be maximized for each radial position $x = \omega r/V_0$.

Doing this, one utilizes following relationship between a and a' , which can be derived geometrically from Figs. 4 and 5 and assuming the induced velocity to be in the same direction as the lift

$$x^2 a'(1+a') = a(1-a) \quad (17)$$

Assuming zero drag and optimizing $f(x)$ and integrating Eq. 16 numerically yield the following curves shown in Fig. 7a for the power coefficient as a function of the tip speed ratio and the number of blades. These curves show the theoretical maximum of C_p when all losses are neglected. It is also seen for a fixed tip speed ratio how the power depends on the number of blades. The asymptotic value for an infinite number of blades and an infinite tip speed ratio can be shown to be $C_{p,\max} = 16/27 \approx 60\%$ and is denoted the Betz limit and is considered the highest possible ratio of the available power that can be harvested from the wind. This occurs at a thrust coefficient, $C_T = 8/9 \approx 0.89$, where C_T is defined as

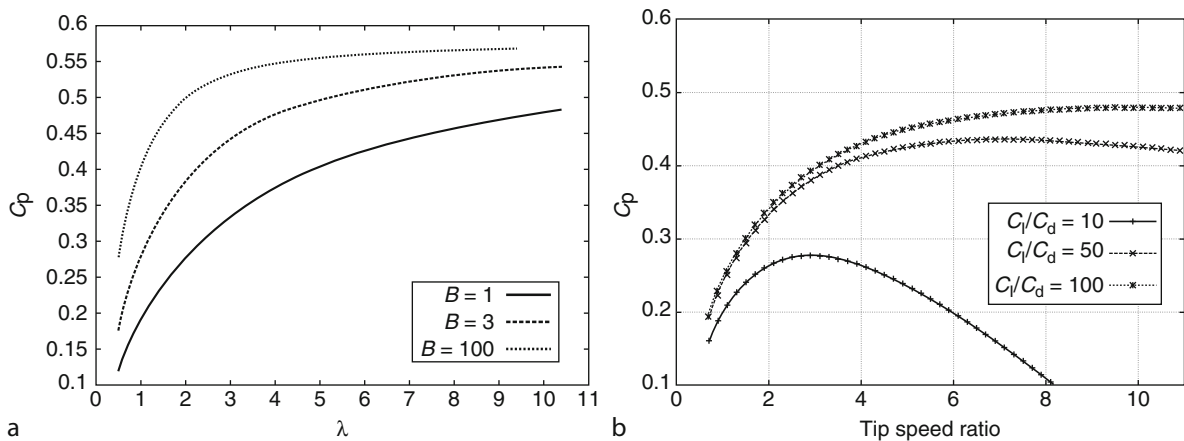
$$C_T = \frac{T}{\frac{1}{2}\rho V_0^2 A} \quad (18)$$

If one reduces the thrust force, T , below this value, the slowing down of the wind speed is reduced and thus

also the decrease in kinetic energy that can be harvested into shaft power. If, on the other hand, the thrust is increased above the optimum value, then air is flowing around the wind turbine instead of through the rotor disc and thus reducing the effective mass flow through the rotor and again reducing the decrease in kinetic energy from upstream to downstream of the wind turbine. In Fig. 7b, it is shown how the power coefficient depends on the tip speed ratio when drag is taken into account and it is seen that to maximize C_p , one should increase the lift to drag ratio.

Environmental Conditions

A wind turbine rotor is not facing a constant wind but a wind field that is continuously changing in time and space due to atmospheric turbulence, wind shear, and influence of the tower. The wind shear is created by the friction between the wind and the ground, and the atmospheric turbulence is produced mechanically and thermally. The thermal contribution comes from an uneven heating of the surface, and if the temperature of the air close to the surface becomes high compared to air above, a plume of air may rise creating a mixing of the flow. The mechanical mechanism is mainly driven by the unstable boundary layer developed by the friction with the surface, where the wind speed becomes zero. Also the flow past obstacles and complex

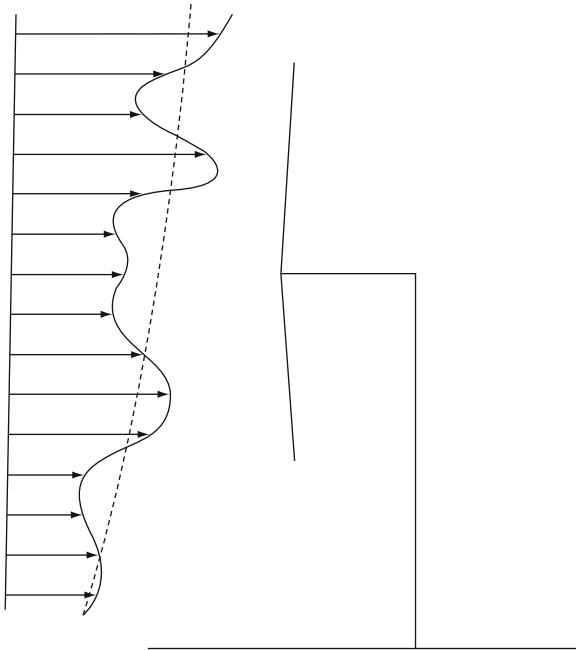


Wind Power, Aerodynamics and Blade Technology. Figure 7

(a) The theoretical maximum power coefficient, neglecting all losses, as a function of the tip speed ratio and number of blades. (b) The theoretical maximum power coefficient, when the drag is also taken into account, as a function of the tip speed ratio for a fixed number of blades of 3

terrain will cause mechanically generated turbulence. Figure 8 shows a snapshot of such an inflow to a wind turbine. The wind shear is drawn by the dashed line that represents the time-averaged velocity at a given height. Another very severe example is when a wind turbine is placed behind one or more wind turbines in a wind farm. In this case, the turbine will experience higher turbulence due to the extra turbulence created by the wake of the upstream turbines.

Since the inflow varies in time and space, a local segment on one of the blades will experience a constantly changing wind speed, and thus the angle of attack and thereby also the aerodynamic loads as indicated in Fig. 4 will constantly change in time and will thus contribute to fatigue damage. So before the loads and thus the output of a wind turbine are calculated, one needs to generate a realistic time and space history of the wind field. Realistic in the sense that it has the same statistical properties as the free wind, both with respect to the power spectrum and spatial correlations as e.g., described in the IEC norm [6]. Some of the most used methods to create a 3D time-varying wind field are described in [7, 8].



Wind Power, Aerodynamics and Blade Technology.

Figure 8

A snapshot of a typical inflow to a wind turbine

Blade Element Momentum Theory, BEM

The most used tool for predicting aerodynamic loads on a wind turbine is the so-called BEM method (Blade Element Momentum method). The BEM model is very popular since it is extremely fast compared to more sophisticated methods such as Computational Fluid Dynamics. A fast method is needed since the loads on a wind turbine are by nature time dependent mostly due to the time-varying incoming wind. Often time series of 600 s are required and a typical time increment, Δt , could be 0.02 s. That is the loads have, for each time series, to be calculated approximately 30.000 times. To do a full design calculation following, e.g., the IEC norm [6] would require in the order of 1,000 such time series. It is therefore in the next many years completely unrealistic to do CFD for a complete design verification.

The BEM model is basically a steady model that computes the flow state where the airflow through the rotor is in equilibrium with the aerodynamic loads along the rotor blades. In order to calculate the unsteady loads, several so-called engineering models must be added. They are called engineering models, since they do not solve directly the physics behind the various phenomena but merely models their effects. The most important one is called dynamic wake or dynamic inflow since it introduces a time delay to the change in the air velocities whenever the aerodynamic loads are changed by, e.g., pitching the blades, see e.g., [2, 3]. The time constant for dynamic inflow/wake is for large wind turbines in the order of 10 s and very important, especially for pitch-regulated wind turbines. Another time delay model called unsteady airfoil data or dynamic stall is also added. A section on a wind turbine blade is constantly experiencing a time-varying angle of attack at a relatively high frequency mainly due to not only turbulence but also wind shear, yaw misalignment, and tower passage. The flow over this airfoil section is thus never in equilibrium with the instantaneous angle of attack, but is lagging behind in time but always chasing the instantaneous equilibrium. The time constant for the unsteady aerodynamics is much lower than the dynamic wake/inflow and in the order of 0.01 s, but including the model is important since depending on the mean angle attack, the amplitude, and frequency, it may enhance stability,

see e.g., [9–11]. Also a model is included that redistributes the induced velocities azimuthally whenever the rotor is not perfectly aligned with the wind. This redistribution yields a restoring yaw moment and is thus important for floating wind turbines with very low torsion stiffness in the tower. The BEM method relies on 2-D airfoil data, $C_l(\alpha)$ and $C_d(\alpha)$, to calculate the lift and drag. ρ denotes the air density and c is the width of the local airfoil and denoted the chord.

$$L = \frac{1}{2} \rho V_{\text{rel}}^2 c C_l(\alpha) \quad (19)$$

and

$$D = \frac{1}{2} \rho V_{\text{rel}}^2 c C_d(\alpha) \quad (20)$$

The relative velocity, $V_{\text{rel}} = |\vec{V}_{\text{rel}}|$, and the angle of attack, α , is computed from Fig. 4 as

$$\alpha = \phi - \theta \quad (21)$$

and

$$\vec{V}_{\text{rel}} = \vec{V}_o + \vec{W} - \vec{V}_{\text{rot}} \quad (22)$$

where $\vec{V}_{\text{rot}}$ is the rotational speed of the blade section and $\vec{W}$ the induced velocity as defined by $\vec{W} = (W_{\text{axial}}, W_{\text{tangential}})$. If the induced velocity is known then the relative velocity and thus the flow angle can be calculated by Eq. 22 and then the angle of attack is given by Eq. 21 and finally the aerodynamic loads from Eqs. 19 and 20. The induced velocity is a direct result of the aerodynamic loads on the incoming air. If one puts up the balance between the local thrust force expressed by the momentum deficit in the wake as given by Eq. 7 and expressed directly from the aerodynamic loads Eq. 4, utilizing that $dT = B p_N dr$, then an expression for the axial induction factor, Eq. 23, can be derived. Similarly an equation for the tangential induction factor, Eq. 24, can be derived through the conservation of angular momentum and the local torque from the blades $dM = B r p_T dr$.

$$a = \frac{1}{\frac{4F \sin^2 \phi}{\sigma C_n} + 1} \quad (23)$$

$$a' = \frac{1}{\frac{4F \sin \phi \cos \phi}{\sigma C_t} - 1} \quad (24)$$

ϕ is the flow angle, F is Prandtl's tip loss correction, $\sigma = Bc/2\pi r$ is called the solidity, and $C_n = p_N / \frac{1}{2} \rho V_{\text{rel}}^2 c$

and $C_t = p_T / \frac{1}{2} \rho V_{\text{rel}}^2 c$ are the normal and tangential force coefficients, respectively. 23 and Equations 24 yield the induced velocity in non-dimensional form and they must in principle be solved iteratively since a and a' depend on ϕ , C_n , and C_t that again depend on a and a' . Equations 23 and 24 express the situation where the induced velocity is in equilibrium with aerodynamic loads. However, the induced velocities change very slowly in time due to the inertia of the mass of air affected by the rotor loads. This phenomenon is denoted dynamic inflow or dynamic wake and a typical time constant is in the order of 10 s. In an unsteady BEM calculation where the time step is in the order of 0.01–0.05 s, the induced velocity changes very little in one time step and one can thus simply use the value from the previous time step on the right hand side of Eqs. 23 and 24 when updating a new set of quasi steady values of a and a' . Having done this, one puts these quasi steady values through a dynamic wake filter to simulate the inertia as, e.g., shown in [2] to determine the final induced velocity at the new time step. Knowing this, one can now compute at this new time the aerodynamic loads as described above and one can then proceed to the next time step. If the thrust coefficient, C_T , as defined in Eq. 18 is large, that is, the force responsible to slow down the wind speed becomes big compared to the momentum of the incoming flow, then the flow through the rotor is no longer comprised by nice streamlines as sketched in Fig. 6 but is much more complicated. This case is denoted the turbulent wake state and is characterized by large zones in the wake containing reversed flow. Here the momentum theory simply breaks down and Eq. 23 is replaced by an empirical relationship between the thrust coefficient and the axial induction factor as

$$C_T = \begin{cases} 4a(1-a)F & a \leq \frac{1}{3} \\ 4a(1 - \frac{1}{4}(5-3a)a)F & a > \frac{1}{3} \end{cases} \quad (25)$$

The BEM is so fast that it on a fast computer can calculate the dynamic behavior of a wind turbine in real time, also taking the structural response of the wind turbine into account. Doing this, the structural velocities are included when estimating the angle of attack, and the strong coupling between the aerodynamic loads and the structural response is denoted aeroelasticity.

Computational Fluid Mechanics

Instead of solving the global momentum equations as in the BEM model one could alternatively solve the Navier-Stokes equations, which are also momentum equations but in differential form. This has the advantage that none of the so-called engineering models are needed, since all the physics modeled through time delays as, e.g., the response of induced velocities on a change in loads is directly contained in the governing flow equations. Also if a full Navier-Stokes simulation is performed, then the uncertainty in the 2-D airfoil data, $C_l(\alpha)$ and $C_d(\alpha)$, is in principle removed. This is of course provided that it is possible to calculate correctly the flow on the blades including the transition from laminar to turbulent boundary layer and the correct separation lines. Full 3-D CFD calculations are made since the 80th about fixed wings for airplanes and helicopter rotors using unsteady Euler solvers [12–15]. The first full Navier-Stokes solutions of wind turbine rotor aerodynamics were published in the 1990s [16–21], where [16] is considered the first complete wind turbine rotor calculation using the 3-D Navier-Stokes equations. However, the first viscous 3-D calculation of a rotating wind turbine blade was made using a viscous-inviscid interaction technique [22], where the potential flow applying a panel method was coupled with the 3-D boundary layer equations. With this model, it is possible to calculate the flow up to a moderate separation. The breakthrough of CFD for wind turbines came with the blind test organized by the National Renewable Energy Laboratory, NREL, after a controlled measurement campaign of a real wind turbine in the NASA Ames 80 × 120 ft wind tunnel [23, 24]. CFD computations made by Risø National Laboratory [25] fitted very well the measured flow and pressure distributions even at high angles of attack and have later inspired other research groups and industry to apply CFD in their aerodynamic research.

Doing CFD one has first to create a computational grid resolving the blades and the outer flow domain (in and outflow). This grid should in principle be fine enough to capture all the scales of the vortices comprised in the flow. It is not possible to resolve the smallest turbulent eddies since this would require an enormous number of gridpoints. Therefore one often

solves the Reynolds Averaged Navier-Stokes, RANS, equations where the equations have been time averaged. To solve these, one needs to model the Reynolds stresses coming from the unsteady perturbations from the turbulent eddies. Most of the recent RANS simulations for wind turbines were done using the $k-\omega$, SST model [26] or Spalart-Allmaras [27]. The RANS models do, however, have some problems computing accurately the flow and thus blade loads in massive stall. It seems that more modern methods like the so-called Detached Eddy Simulation, DES [28, 29], give better results in stall, but at a much higher computational cost. In a DES model, the dynamics of the larger vortices bigger than the computational mesh is directly simulated, whereas the influence of the smaller vortices is modeled similar to RANS. There is hope that when a DES model is used together with a 3-D transition model, very good results can be obtained also at massive stall. But even then there is no way that this type of CFD models can replace the BEM method in design validation calculations were up to thousand time series of 600 s needs to be calculated. But CFD is used to validate the airfoil data used in the BEM calculations for new designs and they are used together with measurements to check and calibrate the engineering models needed by the BEM method.

There exist some hybrid methods that sometimes are used in the design calculations and where CFD replaces the calculation of induced velocities, but where the actual boundary layers on the blades is replaced by volume forces determined from airfoil data in a similar way as in the BEM method. One such model is the Actuator Line model, AL model [30], where the kinematics of the wake is determined directly by Navier-Stokes equations but the aerodynamic loads calculated using tabulated airfoil data and estimating the local angle of attack in the computational grid at the position of the blades. These methods are very well suited to use when calculating the flow within a wind park, where turbines are placed behind other turbines.

Another type of CFD model used for wind turbines is the vortex methods, where vorticity is shed into the wake depending on the time variation of the lift (bound circulation). One such method is the GENUVP code [31].

Rotor Blades

Having determined the outer geometrical blade shape (airfoils, twist, and planform) using, e.g., a BEM code together with some optimization algorithm, one has to build a blade that can withstand the ultimate and fatigue loads. Most wind turbine blades are built from glass fiber mats glued together by epoxy or polyester. To carry the shear forces, so-called shear webs connect the suction side of the blade with the pressure side. These shear webs also ensure the thickness of the blade, that is, that the blade is not squeezed together, when the blade is bent by the aerodynamic loads. When building a blade, one often makes a shell of the pressure and suction side, respectively, and glues the two parts and the shear webs together as sketched in Fig. 9. One has to prove by applying all the load cases described in the norms [6] that the structure will not fail during its design lifetime of typically 20 years.

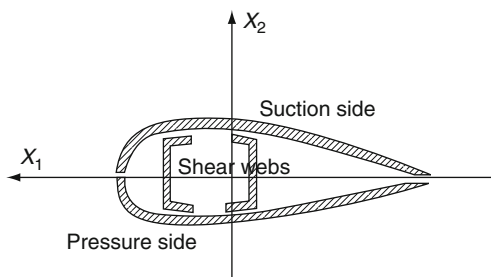
When a blade is geometrically upscaled from one radius, R_1 , to a larger radius, R_2 , all lengths are scaled with the ratio R_2/R_1 , see Fig. 10. By operating the new blade at the same tip speed ratio as the old smaller blade, one ensures that the local blade sections at similar non-dimensional positions, r/R , experience the same flow angle and thus the same angle of attack.

Also the rotational speed is the same at similar non-dimensional positions, r/R , along the blade and thereby also the relative velocity. The only thing changing for the aerodynamic loads, Eqs. 19 and 20, is the chord, which increases as R_2/R_1 . That is the aerodynamic loads increase linearly with the radius, and the blade bending moments, $M = \int r p(r) dr$, thus increase as the third

power of the upscaling. The material stress is calculated about the principal axes, see Fig. 9, as

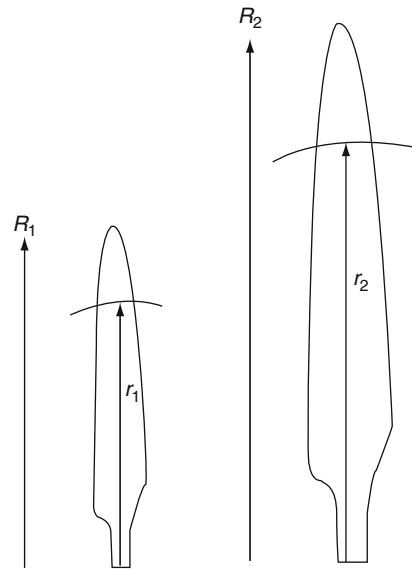
$$\sigma = \frac{M_1}{I_1} x_2 - \frac{M_2}{I_2} x_1 \quad (26)$$

and since the moment of inertia, I , increases with the fourth power of the length, the material stress from the aerodynamic loads is unchanged when upscaling a blade. The minus sign on the second term on the right hand side of Eq. 26 is because compression stress is considered positive and tension negative. For gravity loads, the upscaling causes problems, since the mass per length, m [kg/m], increases with the power of two of the radius, i.e., the bending moments from gravitational loads, $p = mg$, increases as the fourth power with the radius and finally the material stress will increase linearly with the radius according to Eq. 26. This puts a limit to how much one can do a simple upscaling of a wind turbine blade, because at some point, the stress from the gravitational loads will exceed the ultimate stress that the material can carry before failure. So when



Wind Power, Aerodynamics and Blade Technology.
Figure 9

Sketch of a cross section in a typical wind turbine blade



Wind Power, Aerodynamics and Blade Technology.

Figure 10

Geometrically upscaling of a blade multiplying all lengths by R_2/R_1 . Keeping the same tip speed ratio, $\lambda = \omega R/V_0$, the angles of attack at the non-dimensional positions, r/R , are the same

building bigger and bigger blades, one has to make the blades relatively lighter by using less material or at some places changing to a stronger and lighter material, e.g., carbon fiber.

Future Directions

In the future, one will still rely on the BEM method, when calculating all the design cases required by the norms. Probably also the BEM method will be used in optimizations, since this normally requires many calculations of the design which is changing during the optimization process. However, the final design will be checked using full 3-D CFD to see if there is some unexpected flow taking place. Further, CFD will be used to estimate the local lift and drag coefficients along the blade needed as input to the BEM method. In many years to come, the CFD tools will be sharpened with respect to transition models, turbulence models, and computational efficiency/resolution, that is, the affordable number of gridpoints.

Finally, hybrid methods such as the Actuator Line method will be applied when investigating the aeroelastic behavior of wind turbines placed in large wind farms, where the inflow is heavily altered from the upstream turbines compared to the turbines in the first row seeing the undisturbed wind.

Bibliography

1. Glauert H (1935) Airplane propellers, Division L. In: Durand WF (ed) Aerodynamic theory, vol 4. Springer, Berlin, pp 169–360
2. Snel H, Schepers JG (1995) Joint investigation of dynamic inflow effects and implementation of an engineering method. ECN-C-94-107
3. Schepers JG, Snel H (1995) Dynamic inflow: yawed conditions and partial span pitch control. ECN-C-95-056
4. Betz A (1926) Wind energy and its use by windmills. Ökobuch, Staufen
5. Hansen MOL (2008) Aerodynamics of wind turbine, 2nd edn. Earthscan, London
6. IEC 61400-1 Ed.3 CD. 2. revision (2004) 'Wind Turbines. Part 1: Design requirements,' edited by IEC TC88-MT1
7. Veers P (1988) 'Three-dimensional wind simulation' SANDIA report. SAND88-0152 UC-261
8. Mann J (1998) Wind field simulation. *Probl Eng Mech* 13: 269–282
9. Øye S (1991) Dynamic stall, simulated as a time lag of separation. In: McAnulty KF (ed) Proceedings of the 4th IEA symposium on the aerodynamics of wind turbines, Rome, ETSU-N-118
10. Leishman JG, Beddoes TS (1989) A semi-empirical model for dynamic stall. *J Am Helicopter Soc* 34(3):3–17
11. Hansen MH, Gaunaa M, Madsen HA (2004) A Beddoes-Leishman type dynamic stall model in state-space and indicial formulations. Risoe-R-1354(EN)
12. Sankar NL, Wake BE, Lekoudis SG (1986) Solution of the unsteady Euler equations for fixed and rotor wind configurations. *J Aircr* 23(4):283–289
13. Jameson A, Baker TJ (1983) Solution of the Euler equations for complex configurations. AIAA Paper 83–1919
14. Pulliam TH (1984) Euler and thin layer Navier-Stokes codes: ARC2D, ARC3D, notes for computational fluid dynamics users workshop, The University of Tennessee Space Institute, 12–16 March 1984
15. Agarwal RK, Deese JE (1997) Euler calculations for flowfield of a helicopter rotor in hover. *J Aircr* 24(4):231–238
16. Hansen MOL, Michelsen JA, Sørensen NN (1994) Navier-Stokes solver for rotating wing. In: European wind energy conference, Thessaloniki, pp 557–561
17. Hansen MOL, Sørensen JN, Michelsen JA, Sørensen NN (1997) A Global Navier-Stokes Rotor prediction Model. AIAA 97–097
18. Sørensen NN, Hansen MOL (1998) Rotor performance predictions using a Navier-Stokes method. AIAA 98–0025
19. Xu G, Sankar LN (1999) Computational study of horizontal axis wind turbines. AIAA 99–0042, Reno, January 1999
20. Duque EPN, van Dam CP, Hughes S (1999) Navier-Stokes simulations of the NREL combined experiment phase II rotor. In: Proceedings 1999, ASME wind energy symposium, 37th AIAA aerospace science meeting and exhibit, AIAA 99–0037, Reno, January 1999
21. Sørensen NN, Michelsen JA (2000) Aerodynamic predictions for the unsteady aerodynamics experiment phase-II rotor at the National Renewable Energy Laboratory. AIAA-2000-0037
22. Sørensen JN (1986) Three-level viscous-inviscid interaction technique for the prediction of separated flow past rotating wing. PhD thesis, AFM-83-03, Technical University of Denmark
23. Fingersh LJ, Simms D, Hand M, Jager D, Contrell J, Robinson M, Schreck S, Larwood S (2001) Wind tunnel testing of NREL's unsteady aerodynamics experiment. AIAA-2001-0035 Paper, 39th Aerospace sciences meeting & exhibit, Reno
24. Simms D, Schreck S, Hand M, Fingersh LJ (2001) NREL unsteady aerodynamics experiment in the NASA-Ames wind tunnel: a comparison of predictions to measurements. NREL/TP-500-29494, June 2001
25. Sørensen NN, Michelsen JA, Schreck S (2002) Navier-Stokes predictions of the NREL Phase VI rotor in the NASA Ames 80-by-120 wind tunnel. AIAA-2002-0032
26. Menter FM (1993) Zonal two equation $k-\omega$ turbulence models for aerodynamic flows. AIAA-paper-932906
27. Spalart PR, Allmaras SR (1992) A one-equation turbulence model for aerodynamic flows. AIAA-92-0439
28. Travin A, Shur M, Strelets M (1999) Detached-eddy simulation past a circular cylinder. *Flow Turbul Combust* 63:293–313

29. Strelets M (2001) Detached eddy simulations of massively separated flows. AIAA-2001-0879
30. Sørensen JN, Shen WZ (2002) Numerical modelling of wind turbine wakes. *J Fluids Eng* 124(2):393–399
31. Voutsinas S, Belessis M, Huberson S (1993) Dynamic inflow effects and vortex particle methods. In: *Proceedings of the European community wind energy conference, Germany*, pp 428–431

Wind Power, Introduction

LENNART SÖDER

Royal Institute of Technology, Electric Power Systems,
Stockholm, Sweden

Article Outline

Meteorology and Wind Power
 Aerodynamics and Blade Technology for Wind Power
 Wind Power Grid Integration: Local Grid, Power
 Quality and Wind Power Generator Systems
 Offshore Wind Power, Including Connection to
 Shore
 Wind Power, Balancing of All Sustaining Factors
 Wind Power: Economy, Market, Subsidies, Payment
 Mechanisms, and Capacity Credit
 Wind Power: Transmission Planning
 Wind Power Systems, Small Scale
 Wind Power – Environmental Impact: Noise Emissions
 Wind Power: Basic Challenge Concerning Social
 Acceptance
 Wind Power Installations, Global
 Bibliography

Windmills have existed for at least 3,000 years, mainly for grinding grain or pumping water. The use of windmills to generate electricity is a twentieth-century development and in the early 1970s started the development toward modern wind turbine technology. Wind power has increased significantly during the last decade and there are currently many power systems, or subsystems, with comparatively large amounts of wind power. In 2009, Spain covered 14% [1] of their electric energy demand with wind power. The corresponding figures were for Portugal 15% [2], Ireland 11% [3], and Western Denmark 25% [4].

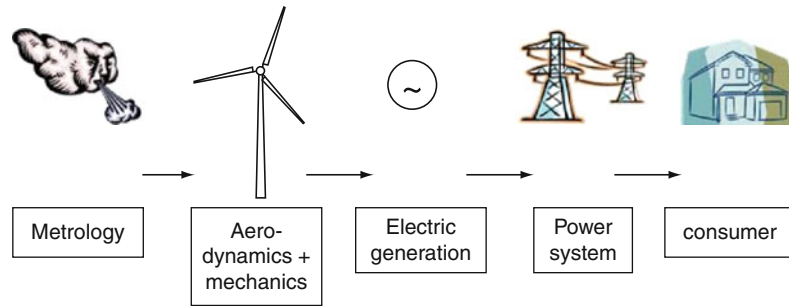
This means that wind power has developed from a marginal source to a source with important impact on national power system balances. But in many countries, wind power still only contributes to a small part of the electric supply. The total installed capacity at the end of 2009 reached 158,000 MW including 37,000 MW installed during 2009. A third of these additions were made in China, which experienced yet another year of over 100% growth [5]. From 1996 to 2009 the mean increase of the amount of wind power was 28% per year.

There are several reasons why wind power increases so fast. They include (a) more cost-efficient technology, (b) higher energy prices depending on high demand and decreasing oil resources, and (c) increasing interest in energy sources with low emissions of CO₂ and other gases.

To be able to develop this type of system it is necessary to combine knowledge from several areas such as meteorology, aerodynamics, mechanics, electric machinery, and power systems. In order to obtain a cost-efficient wind power system with low environmental impact, all these parts have to be integrated together with a power system which can balance and transfer the produced power from the wind power plants to the power consumers.

Figure 1 shows the basic energy flow from the wind to electricity for the final consumer.

The basic energy flow starts with the wind. Winds are movements of air masses in relation to the ground surface. The winds are caused by pressure forces and sometimes also by the gravitation. The main origin of winds is temperature differences. The horizontal component is dominating. The energy in the wind comes from the mass, that is, the air, which is in motion. This is called the kinetic energy of the wind. To “use wind energy” is the same as transferring this kinetic energy into, for example, electric energy. Thereby the kinetic energy is decreased, that is, the wind speed is decreased. A wind turbine transfers the energy in the wind into a turning torque on the turbine shaft. A basic understanding of how this works can be obtained from aerodynamics and classic theory concerning forces when an airfoil moves in a flow. The turning torque of the turbine shaft then drives the generator and the output is then electric energy, that is, the generator transfers the mechanical energy in the rotating shaft into electric energy out from the



Wind Power, Introduction. Figure 1

Energy transfer from the wind resource to the final consumer

generator. The electric energy is then transferred to the consumers on electric wires and consumers finally use the energy for lightning, heating and cooling or for appliances such as computers or other electric machines. There must also be a system solution in order to supply the consumers with electricity exactly when they need it, which may not be the same situation as when the wind blows.

This section on Wind Power is comprised of 12 detailed entries which present the technology basis of this important source of electric power. The chapter on “[Meteorology and Wind Power](#)” describes the development in Wind Power Meteorology followed by “[Aerodynamics and Blade Technology for Wind Power](#),” which contains a description of the function and efficiency of a blade. The generation of electricity is then described in “[Wind Power Grid Integration: Local Grid and Power Quality and Wind Power Generator Systems](#)” and the specific challenges of offshore installations are shown in “[Offshore Wind Power, Including Connection to Shore](#).” Since wind power is changing in intensity, the balancing of this source is crucial which is described in “[Wind Power, Balancing of all Sustaining Factors](#)” while in “[Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit](#),” the economics of wind power is handled. Transmission issue of wind power grid integration is discussed in an interesting contribution called “[Wind Power: Transmission Planning](#).” Another interesting and increasing possibility is small-scale wind power which is treated in “[Wind Power Systems, Small Scale](#).” The environmental impact and acceptance are handled in “[Wind Power – Environmental Impact: Noise Emissions](#)” concerning noise emission

and in “[Wind Power: Basic Challenge Concerning Social Acceptance](#)” with the issue of public acceptance. The technology, economy, and acceptance have led to a global fast increase of wind power which is finally summarized in “[Wind Power Installations, Global](#).” Key aspects of each of the above listed entries are presented below.

Meteorology and Wind Power

The availability of a suitable wind resource at a turbine site and the ability to accurately calculate it over the 20–year life span of the turbine are essential preconditions for the developments of wind energy as a major global provider of electricity. Therefore, the objective for research on wind conditions for wind energy resource and design purposes is to be able to determine the wind conditions with a high accuracy at any place on the Earth which could be a potential site for wind energy exploitation. The wind conditions at a specific site are basically dependent on two factors: the regional climatology and the local topography. The current fast development in Wind Power Meteorology can be termed “From Global to Local” which is synonymous with the process that goes from catching the relevant information from Global dataset, applying the data in meteorological models from global to mesoscale, which in turn produces spatial and temporal information to be used by microscale models to produce the relevant information on the wind conditions for design and for power production at a specific site for specific wind turbines. Therefore, most studies of wind conditions for wind power still rely on the use of local measurements in combination with microscale models.

Aerodynamics and Blade Technology for Wind Power

The energy extracted from a wind turbine comes from extracting some of the kinetic energy contained in the wind. Assume that the wind speed is constant, and that the wind turbine is rotating at a constant speed. This implies that each part of the blade is hit by a constant wind speed, but not the same total wind speed since this is the vector sum of the wind speed and the turning speed, which increases depending on the distance from the center of rotation, since the rotational speed increases proportional to the radius. The wind speed and the turning speed are about perpendicular since the wind turbine is turned toward the wind. The airfoil has an angle of attack of α to the air stream. This angle is the angle between the chord, c , and the direction of the flow, V . This angle then changes if only wind speed changes, only rotational speed changes, or if the blade is pitched.

The aerodynamic efficiency is connected to the ratio between the tip speed and the wind speed. From efficiency point of view, the blades should move faster, that is, higher speed of the shaft, when the wind speed is higher, and this has an impact on the requirements of the electric generation system.

Since the wind speed varies in time and space, a local segment on one of the blades will experience a constantly changing wind speed, and thus, the angle of attack and, thereby, also the aerodynamic loads will constantly change in time and will thus contribute to fatigue damage. This means that the amount of turbulence in the wind has to be considered when blades are designed. Having determined the outer geometrical blade shape (airfoils, twist, and platform) using, for example, the Blade Element Momentum theory (BEM), together with some optimization algorithm one has to build a rotor blade that can withstand the ultimate and fatigue loads. Most wind turbine blades are today built from glass fiber mats glued together by epoxy or polyester.

Wind Power Grid Integration: Local Grid, Power Quality and Wind Power Generator Systems

The electrical system of a wind turbine could be referred to as the heart of the wind turbine. The main purpose of the electrical system is to convert the

incoming mechanical energy on the shaft to electrical energy that is fed into the connected power grid. This conversion should be done with a minimum of losses as well as with the smallest size possible.

In addition, the electrical system has the ability to influence the power quality impact that a wind turbine has on the power grid and also the ability to control the speed and torque on the machine shaft (if it is equipped with a power electronic converter) and, in this way, to maximize the energy capture as well as to keep the stresses on the structure to a minimum.

These requirements (high energy capture, good power quality impact, low stresses on the structure in order to minimize the material needed) have in combination with the technical development led to an evolution of wind turbines and definitely regarding the electrical systems of wind turbines. This development is continuously ongoing and is the reason to why there are several different concepts of electrical generating systems for wind turbines.

Offshore Wind Power, Including Connection to Shore

Offshore Wind Power is defined as construction of wind farms, located in the water, generating electricity from wind. Land-based wind power will remain dominant in the immediate future but installations at sea will become increasingly important. There are already today, 2010, several installations in Denmark, Sweden, UK, and Germany.

Compared to onshore wind, offshore wind is more complex and costly to install and maintain but also has a number of key advantages. Winds are typically stronger and more stable at sea than on land, resulting in significantly higher production per unit installed. At sea, wind turbines can be bigger than on land because of the logistical difficulties of transporting very large turbine components from the place of manufacturing by road to installation sites on land.

Wind Power, Balancing of All Sustaining Factors

Wind power balancing considers the challenge to keep a continuous balance between total production and total consumption in a power system. The varying wind speed introduces special considerations

concerning how to keep this balance in an efficient way. In a power system, there is always a challenge to keep the balance between total production and total consumption. This challenge has existed ever since the first power systems were installed more than 100 years ago. With the increasing amount of wind power new challenges arise.

There are always changes in production and consumption in all power systems since the load changes continuously and there are sometimes sudden outages in power plants and/or power lines. Despite these changes, the continuous balance between production and consumption is always kept. With larger amounts of wind power, the net changes in the controllable part of the production and consumption system will increase since wind power changes will increase the overall changes seen by the rest of the system. The challenges include balancing of changes, and keeping enough margins for reacting to changes depending on uncertain forecasts. Also, methods for estimation of the size of the increased challenges need to be addressed, as well as efficient market rules for an economic operation. When the share of wind power increases or is expected to increase it is common that so-called “integration studies” have been performed and the setup of these ones can be essential for a realistic estimation of the consequence of more wind power.

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit

With increasing concern regarding climate change and emissions from conventional power plants, new markets have developed that allow for the trading of the clean energy attributes of wind and other renewable energy sources. Two examples are green power markets and renewable energy standards. Wind energy can also be bought and sold in wholesale energy markets including bulk power exchanges. In some areas, capacity markets exist that provide payment for installed capacity that is available when needed. Capacity markets are generally based on the capacity contribution that a power plant or responsive load can make toward planning reserves, which is called the capacity credit or capacity value. The general

motivation behind a capacity market is to ensure sufficient revenue for the generator and ensure sufficient installed capacity.

In some countries, government incentives exist that encourage the development and use of wind energy (or other renewable energy sources). These include, for example, renewable energy standards, wind energy tax credits, or green power markets. Because of its variable and uncertain characteristic, wind energy can increase overall operating reserves that are needed to manage uncertainty. Specific methods to measure this cost have evolved, and today, there is some controversy surrounding the topic and its precise definition. Integration impacts and costs are not unique to wind. Nuclear power, for example, typically requires larger contingency reserves and increases cycling of other base-load generation. All investments in the power system impact system operations and costs.

Wind Power: Transmission Planning

The transmission grid is essential for an efficient use of large-scale introduction of wind power. With the transmission grid it is possible to, to some extent, even out the variation of wind power in large areas and to coordinate the balancing resources in the different interconnected areas. Transmission planning includes both the internal planning in each subregion and the overall planning for the whole interconnected system. Overlay transmission systems can be designed to economically and reliably serve load and deliver renewable energy. Overlays are a layer of transmission which spans the underlying transmission systems. Overlays require different planning techniques than those of the underlying systems. The involvement of many different people with widely varying backgrounds and interests have driven the transmission planning processes to provide more information than has been required in the past. Larger transmission expansions have attracted more stakeholders in the process.

Wind Power Systems, Small Scale

There is a growing interest for the use of small wind systems. Small wind systems can be seen in autonomous operation for pumping water, desalinization, and

battery charging. A growing amount of small wind turbines is connected to the grid, most of them behind the meter. Electricity production with wind turbines on buildings is a fairly new development with an interesting potential. Safe and reliable operation of small wind turbines is technically challenging since they usually operate in more turbulent flows than large-scale wind turbines. For an efficient application of small-scale wind power systems it is important to have knowledge of working principles of various type of the technology, their safety and control system, noise and vibration as well as investments costs and the cost of energy.

Wind Power – Environmental Impact: Noise Emissions

Wind turbines generate sound. Mainly this sound source is of aeroacoustic origin as improved sound insulation of the nacelles has reduced the mechanical sound sources. Unwanted sound is called noise and is a major factor when assessing nuisance caused by wind turbines. The increasing height of wind turbines has led to better possibilities to harvest wind at higher altitudes but has also contributed to increased noise levels in stable atmospheric conditions when the sound is easily transmitted due to a high wind speed gradient. Rapid development of large wind farms in rural, earlier unexploited, areas could cause large acceptance problems for nearby residents if noise is not properly considered in the development process.

Wind Power: Basic Challenge Concerning Social Acceptance

Successful implementation of new technologies requires social acceptance. Historically, for the implementation of wind energy this was considered a relatively simple issue that could be addressed by applying effective communication strategies. Without much study, social acceptance of wind power was considered a matter of merely public acceptance, and any problems with public acceptance were viewed as issues of education. Furthermore, public acceptance was primarily negatively defined, under the heading of “nontechnical factors.” Three distinguishing

dimensions of the social acceptance of renewable energy innovations have been recognized, all with its own character and dynamics: sociopolitical acceptance, with a focus on decisions that create favorable conditions for the other two forms of acceptance; community acceptance that concerns the decisions about the integration of wind power scheme at a certain location; and market acceptance, which is about willingness to invest among all different kinds of actors, including willingness to pay for wind-generated power among consumers.

Wind Power Installations, Global

Wind power has been utilized for over 3,000 years to aid with human activities. Although it was predominantly used as mechanical power at the beginning, gradually, with industrialization, the focus shifted to the use as electrical power. With the arrival of the oil price shock in the early 1970s, the use of wind energy as electrical power started to gain more focus, and since the end of the twentieth century, wind energy has become one of the most important sustainable energy resources. The installation of wind power production capacity is growing at a rapid pace, doubling every 3 years, and is expected to continue growing in the future, providing around 20% of the worlds’ power needs by 2030. With strong growth in onshore and offshore developments in Europe, and rapid market expansions in the rest of the world, it is imperative that the knowledge which is currently concentrated in a few countries be spread, as implications on research, education, and electric utilities all around the globe are significant.

Bibliography

1. Spain: http://www.ree.es/ingles/sistema_electrico/pdf/infosis/sintesis_REE_2009_eng.pdf
2. Portugal: <http://www.centrodeinformacao.ren.pt/EN/InformacaoTecnica/TechnicalData/2009.pdf>
3. Ireland: <http://www.eirgrid.com/media/Annual%20Report%202009.pdf> and http://www.ieawind.org/AnnualReports_PDF/2009/Ireland.pdf
4. Denmark: http://www.danskeenergi.dk/~media/Energi_i_tal/Dansk_Elforsyning_Statistik_2009.pdf.ashx
5. Global Wind Energy Council (2008/2009) Global installed wind power capacity (MW). <http://www.ewea.org>

Wind Power: Basic Challenge Concerning Social Acceptance

MAARTEN WOLSINK

Department of Geography, Planning and International Development Studies, University of Amsterdam, Amsterdam, The Netherlands

Article Outline

Glossary

Definition of the Subject

Introduction: Four Starting Points

Dimensions of Social Acceptance of Energy Innovation Actors

Subjects of Acceptance

Sociopolitical Acceptance

Community Acceptance

Market Acceptance

Significant Attributes Connected to Identity

Future Directions

Bibliography

Glossary

Attitude Disposition to evaluate a psychological object – the attitude object –, representing a summary evaluation of this object captured in such attribute dimensions as good–bad, harmful–beneficial, pleasant–unpleasant, and likable–dislikable.

Attribute Character ascribed to an attitude object about which an actor may hold a belief (expectancy) and an evaluation (value).

Collaborative planning Planning with delegated responsibility to stakeholders who engage in interest-based negotiation about a plan or a project.

Community A body of people viewed collectively, e.g., the local community surrounding a wind farm location or a community holding a collective interest.

Discourse A shared way of apprehending the world; in this case, reflecting how the environment (including wind power implementation) is interpreted and given meaning.

Framing The way an issue (or an attribute) is defined and presented by actors – biased by their own perspective – in order to affect the perception of

the issue by others: to encourage certain interpretations and to discourage others.

Innovation A change of ideas, that becomes manifest in products, processes, or organizations, that are applied successfully in practice.

Institutions Existing patterns of behavior, determined by existing societal rules “the rules of the game in a society.”

Landscape The part of the environment that is the human habitat as it is perceived and understood through the medium of our perceptions.

NIMBY Depreciative interpretation and characterization of opposition to a facility: an attitude of objection to the siting of a facility in the proximity (“backyard”), while by implication raising no such objections to similar developments elsewhere; acronym of “not-in-my-back-yard.”

Place identity Human binding to the physical environment at a certain place or area associated concepts: place attachment, sense of place.

Public acceptance The degree to which a phenomenon is taken by the general public, the degree to which the phenomenon is liked by individual citizens.

REFIT Renewable energy feed-in tariff, a class of financial procurement systems creating a priority market for renewable generated electricity by guaranteed access to the grid with a long-term fixed price per kilowatt-hour.

RPS Renewable Portfolio Standards, a class of financial procurement systems based on certificates issued for renewable generated electricity – “green certificates” – with a legal quote for renewables creating a market for trading certificates.

Smart grid Power grid consisting of a network of integrated microgrids that can monitor and heal itself.

Social acceptance The degree of which a phenomenon (e.g., wind power implementation) is taken by relevant social actors, based on the degree how the phenomenon is (dis-)liked by these actors.

Socio-technical system A system be made up of scientific and technological, as well as socio-economic and organizational components.

Definition of the Subject

Successful implementation of new technologies requires social acceptance. Historically, for the implementation

of wind energy this was considered a relatively simple issue that could be addressed by applying effective communication strategies. Without much study, social acceptance of wind power was considered a matter of merely public acceptance and any problems with public acceptance were viewed as issues of education. Furthermore, public acceptance was primarily negatively defined, under the heading of “nontechnical factors” [1].

However, innovation must be considered a much broader concept, as there is nothing inevitable about how new technology is developed and implemented. The willingness to accept phenomena related to innovation of different parts of society, including all realms beyond “the public,” can be subdivided in two broad categories:

- Acceptance of the creation of new socioeconomic conditions needed for implementation.
- Acceptance of the consequences of the implementation: implementation will affect current practices in society and forcing some to change.

The very limited number of first studies on acceptance in the 1980s indeed focused upon public attitudes [1–4]. One of the first observations also was that social acceptance studies should look at the conditions that determine the effective support that applications of wind power would get at all different levels of decision making. This is already a very important conclusion. Actors involved in developers tend to perceive the issue from their own perspective: they only see the resistance they are confronted with against their projects, and they tend to frame this phenomenon as resistance against wind power per se. This is certainly not a phenomenon restricted to wind power, it can be found in almost any domain of infrastructure decision making (roads, waste facilities, power transmission lines, nuclear power facilities, prisons, etc, but also new housing districts or schools). However, if this kind of framing becomes persistent, it might become an even stronger impediment for positive decision about wind power developments, as it obstructs collaborative ways of decision making (see sections “[Mind the “Gap”](#)” and “[The Backyard](#)”).

The initial studies immediately showed that neither public support, nor support from crucial stakeholders at varying scale levels could be taken for granted. The pioneer of social acceptance studies, Carlman, observed

that building wind turbines was “a matter of public, political, and regulatory acceptance,” and therefore she carried out a study on the acceptance of wind power among decision makers [5, p. 339]. Her study suggested that there were several constraints for social acceptance beyond public acceptance.

Soon in the 1980s, other academics joined this approach and specified the definition and analysis of the problematic issues for implementation [4, 6]. These studies focused on issues such as the lack of support among key stakeholders – including the limited commitment of energy utilities – reluctance among policy makers to dedicate themselves to convincing policies, and the lack of understanding of the roots of attitudes toward wind power schemes among local residents and local communities.

Because implementation issues are much broader than public opinion with regards the application of the energy source “wind,” the nature of the studies has changed. The number of studies remained limited during the nineties, but a significant increase took place during the last decade. Many of these studies focused upon the level of decision making about building wind farms, which is the crucial factor in the success rate of a country’s wind power development. As the character of all these cases of building wind farms shows a wide variety, most of these studies are designed as single case studies. Beside those studies, some comparative case studies have been carried out to compare the differences between countries, and single cases that analyzed the development of policies within a country.

The intensive study of social acceptance has a rather short history, and because the concept is complex our understanding of the issue is still limited [7]. Only recently, a first effort to elaborate the concept of social acceptance of renewable energy innovation was made [8]. Three distinguishing dimensions of the social acceptance of renewable energy innovations are now recognized, all with it’s own character and dynamics: sociopolitical acceptance, with a focus on decisions that create favorable conditions for the other two forms of acceptance; community acceptance that concerns the decisions about the integration of wind power scheme at a certain location; and market acceptance, which is about willingness to invest among all different kinds of actors, including wiliness to pay for wind-generated power among consumers.

Introduction: Four Starting Points

Socio-technical Systems

Successful implementation of new technologies requires sociopolitical acceptance. The implementation of wind power is a form of innovation, as it implies the application a relatively new form of energy supply. By definition, innovation is the successful application of new ideas, transferred into products, processes, or organizations. The integration of new technologies such as wind power in society is not a matter of simply applying science; also there is a series of factors in the social environment that co-determine it: social, economic, cultural, and different technological factors [9]. Regarding wind power, these include the integration of this new way of generating power in the existing energy supply and consumption, as well as the appearance of modern wind turbines in our landscapes and as part of communities and local or regional economies.

In addition to technological concerns, the application of wind energy carries a strong socioeconomic component. Significant scientific and technological challenges need to be overcome, but as history shows, well-developed science and technology are not sufficient to foster application of technologies. Making it affordable does not address the entire challenge; the society may fail to allot acceptable locations for the facilities or fail to deliver the required social and financial investments. This observation leads to *basic assumption I*:

- An energy system containing a substantial part of wind-generated power and also a substantial part of wind-generated energy consumption, should be considered a new socio-technical system (STS).

Acceptance of Distinguishable Aspects

A STS is a system with new scientific and technological as well as socioeconomic and organizational components, which is reflecting new ideas and concepts on the proffered design of such new systems [10]. Other related examples of innovations are, for example, the introduction of a new production process in a factory, (an STS combining new techniques with a different organization of the process) limiting the amount of emissions and final waste after consumption. Or the

development and introduction of an ecological certificate (including the organization and techniques needed to generate reliable information, monitoring producers in all phases, and effectively informing potential consumers) aiming to affect consumer choices that would put pressure on producers and distributors in the production–consumption chain. The introduction of electric vehicles also heavily depends upon the organization and implementation of infrastructure to load these cars, including the question who will provide this infrastructure. In fact, the latter example will soon be very relevant for further implementation of renewable energy as well (see section “[Future Directions](#)”).

Any STS pays attention to the linkages between these two components, links that are necessary to fulfil societal functions [11]. Such functions can be – among others – support of regional economies, provision of energy, reduction of environmental impact, etc. The development of distinguishable sub-functions of the new crucial technologies requires favorable institutional conditions and extensive social learning. This also holds for wind power [12, 13]. Hence, the second *basic assumption (II)* states two things:

- All aspects of a new STS featuring a substantial amount of wind power are subject to social acceptance.
- An actor may accept certain aspects, while simultaneously rejecting other aspects, as a result of social, economic, and/or political learning processes.

In all domains, to-be-implemented innovations require institutional changes and careful strategic governance, especially when such innovations concern fundamental shifts in ways of thinking about the way the system is designed and how it should work [10].

Institutional Change

For wind power, it is not only the technical novelties that are subject of acceptance, but rather the new ways of organizing the STS and how to take decisions about it that are required to build the turbines and to integrate their variable energy production into the system of energy supply and demand. In particular, those new ways of organizing require new ways of thinking, which only emerge through processes of organizational and policy learning. Developing an

STS requires the understanding of and the will to change crucial “rules of the game in a society,” the short definition of *institutions* [14]. Institutions are defined as mutually reinforcing patterns of behavior and thinking (of actors of all levels), as reflected in formal and informal rules, norms, and procedures. These patterns of thinking and behavior can be recognized within all realms of society, including governance systems.

The way existing supply and demand of electrical power are shaped is full of patterns of behavior of all kinds of actors. These behavioral patterns are based on formal and informal rules that have emerged over time. These rules are not necessarily defined to further the integration of new energy sources, as they have emerged in the past under different conditions, serving different ends. The most essential changes in ways of thinking concern modes of thought that are historically rooted in the competent organizations. This phenomenon is called “path dependency” and reflects the historical roots of existing institutions [14, 15]. The process “by which actions are repeated and given similar meaning by others” is *institutionalization* [16, p. 117]. The existing ways of things are done and how they are organized all result from past institutionalization processes. That includes, for example, how infrastructure is built and how supply of electricity is organized in the power supply sector. However, at that time, these institutions were framed to serve societal needs that did not include the implementation of a resource like wind, and within a context of a different socioeconomic environment. The existing patterns were not designed with the new STS in mind and therefore path dependency in an existing STS is often responsible for unfavorable conditions that forestall the introduction of a new STS, such as wind power implementation. This may eventually lead to a deadlock in the development of the new system, known as institutional “lock-in.” This is an important source of nonacceptance among different actors with regard numerous elements of wind power implementation. Therefore, the third *basic assumption III* is:

- Existing institutions (existing patterns of behavior as determined by existing societal rules) often impede the development and implementation of new views, approaches, techniques, and practices required for the implementation of wind power.

The social acceptance of wind power is a complex issue due to a combination of two factors: an STS that includes wind power may induce different reactions from an actor (assumption II) and there are many different types of actors who are involved at various scale levels, all with their own institutionally embedded behavioral patterns (assumption III). This complexity provides for many opportunities for misunderstandings. These have emerged in “common sense” approaches with regards learning and acceptance. What is considered “common sense” is often no more than a habitual pattern of thinking that is mutually copied by actors and reinforced by frequent use. This manner of institutionalization, the tendency to copy or follow others in the organization of an actors’ own behavior – known as “mimetic” repetition [17] –, is a strategy used by social actors to address uncertainty as well as to mask the appearance of their own uncertainty. Existing ideas about how to implement wind power and how to integrate it in society are full of such institutionalized ways of thinking, which are often classified as “knowledge.” However, this is often unjustified as they are in fact no more than institutionalized commonsense ideas that may become barriers to actual implementation.

Generalizing, we can recognize the three major commonsense approaches to wind power implementation. These can be found among many different actors at all scale levels, but most relevant is that they belong to policy beliefs of actors (public agencies, governmental, but also stakeholders) that often hold dominant positions in policy. These three commonsense approaches are:

- The so-called “technological fix”: a tendency to focus on technology while not recognizing the social components of the STS. An example with regards wind is the focus on the technological potential of wind power.
- The simplification and reduction of the issue of social acceptance to mere public acceptance. In many policy documents, public acceptance is simply used as a proxy for social acceptance, but without any justification. This is tempting, as it reduced a very complex phenomenon to simple figures of public opinion, but it also reduces the validity of these figures to approximately zero.

- The one-sided focus on the objectors of wind development. In most documents about social acceptance the support for wind power and wind power projects among all different kinds of actors is taken for granted. As a result, it is hardly investigated, hardly valued, and not well understood.

Beyond Common Sense

The first commonsense approach is to simply neglect the issue of social acceptance or to handle it as a secondary aspect of residual questions. From the beginning of modern wind energy development, social acceptance issues were considered elusive and hard to calculate (“soft factors”) and usually negatively defined as “nontechnical factors” [1]. Moreover, these factors were narrowly defined as the term public acceptance was considered synonym to social acceptance. The institutionally determined approach of society’s abilities to implement wind power tends to focus on the so-called “technical potential,” sometimes combined with assessments of the financial feasibility of that potential [18]. Even if social constraints are included in the assessment of the potential, they remain limited to “planning” consideration, based on a uniform criterion for the possibility of land use for the construction of wind turbines [19]. The concept of the technical potential of renewable energy options implies that technology need only be developed and applied to solve all of a country’s energy problems. A study conducted in the USA found that this has become somewhat of a mantra in discussions about both electricity and energy policy more broadly [20, p. 4511]. However, much of the technically and meteorologically available possibilities cannot be utilized without violating generally applied social norms or without challenging existing interests. The social potential is ultimately the potential that really indicates what capacity can be implemented. The main bias in the approach of the technical potential is that it narrowly frames the concept of social acceptance as a collection of “barriers” to achieving our potential.

The combination of the second and third general commonsense approaches has prevented the understanding of social acceptance of wind energy innovation for a long time. Unfortunately, from the viewpoint of wind power implementation, they are

still widespread among developers and policy makers alike [21, 22]. The second commonsense bias is the simplified idea that social acceptance is equal to public acceptance. The reduction of acceptance to public opinion is a strong bias, because it completely ignores the position of all societal actors beyond the level of the individual. The perspectives of these other actors, companies, public bodies, organizations, etc. are certainly not mere translations of personal attitudes, and particularly not of ordinary citizens, and in many cases their influence on decision making and eventually on actual implementation rates is much stronger than that of the public [23].

The third commonsense bias is the one-sided focus on the potential objectors of wind energy development and neglect of the supporting side that accepts such development. Still many studies only focus upon why people oppose and why stakeholders object (e.g., [20]), whereas it is equally important why wind power is supported in the first place [7, 8]. How much support is there – also for wind power schemes, as is in most cases supporters outnumber objectors – and what are the reasons for actors to be in favor? Acceptance is, by definition, taking what is offered, whether by favor or reception. Of course, this definition also implies the inclusion of the side of refusal in the concept. Social acceptance is a bipolar phenomenon. Following definitions from psychology on the social acceptance of individuals [24], the social acceptance of a phenomenon like the implementation of wind power is the degree to which people like or dislike the phenomenon. The concept includes all degrees, from full refusal to total adoption. Therefore, the narrow focus on the opposing forces to wind power developments neglects the essential side of acceptance issues. The bias is that the position of support is apparently taken for granted and considered as the “natural” position. Consequently, the supportive position does not seem to need an explanation, whereas the position of refusal becomes deviant. Understanding social acceptance requires both positions to be analyzed as equally legitimate in principal [7, 23].

Hence, the fourth fundamental proposition for understanding social acceptance is – *basic assumption IV* –:

- The objection to any wind power development must be considered as a potentially legitimate, rational, and

informed position: All positive, active support as well as passive supporting attitudes, are equally important for obtaining a good understanding of the acceptance of wind power.

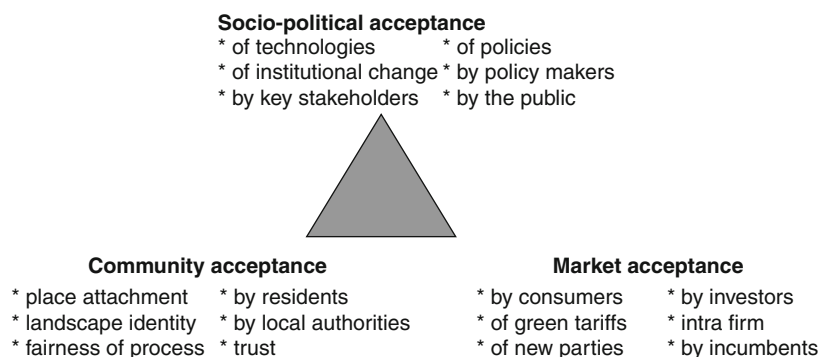
Dimensions of Social Acceptance of Energy Innovation

Although used often, the term social acceptance is seldom properly defined. Here a systematic elaboration is provided, starting from three distinguishing dimensions of the social acceptance of renewable energy innovations [8]. These are presented in Fig. 1, along with the main issues that should be addressed within those dimensions. Sociopolitical acceptance is related to the support or resistance toward policies that affectively promote the implementation of wind power. This is not only related to government. Today “governance” is the preferred concept, which implies that the drafting and implementation of policy is not restricted to actors within the realm of government, but rather it includes many social actors. These so-called “stakeholders” include all social actors who are involved in negotiating, influencing, and lobbying, because they perceive a specific policy as affecting their (vested) interests. The recognition that policy is in fact shaped within a broader arena that includes both private and public actors is referred to as “governance” [25].

Policies tend to focus primarily on the economic investments in new STS. Policy instruments are usually designed (or perceived as designed) to create incentives for positive investment decisions. They are determined

to affect the market decisions of actors. These actors primarily are incumbent energy companies, but they may also include newly emerging actors who are entering the market, or existing actors in other domains who discover wind power as a new and promising investment area. The supporting side of “market acceptance,” which is a willingness to support or take part in investment, is not the only required and favorable position, however. The nature of wind power installations is fundamentally that they are part of the phenomenon of “distributed generation,” in which the power scale, the power delivery (directly to the consumer or to the grid), the mode of operation, ownership, the purpose, and the location of the power generation are as essential as the generated power itself [26]. All these elements are subject to the acceptance within the community in which the generation takes place.

In the commonsense approaches, community acceptance is seen as the bottleneck of wind power development. Indeed, the most problematic issues with regard to acceptance manifest themselves in the discussions about establishing wind farms. By definition, this is a discussion at the level of community acceptance, but the idea that the roots of problematic acceptance lie within the community only emerges because of the combination of two commonsense views: the dominant narrow interpretation of social acceptance as public acceptance and the focus upon objectors while neglecting positive support. The three-dimensions model of acceptance of renewable energy innovation (Fig. 1) shows that social acceptance is much more than simple public opposition on the local level. Before discussing these three



Wind Power: Basic Challenge Concerning Social Acceptance. Figure 1

Three dimensions of social acceptance of renewable energy. Adapted from [8, p. 2684]

dimensions, we focus upon two basic features determining the construction of social acceptance within all three dimensions:

- The different actors: who is taking a favorable or opposing position?
- The different subjects: to what attitude object are those actors defining their position?

Actors

As the first component in the social dimension of the socio-technical system, all relevant social actors involved in the process of acceptance must be considered. Table 1 shows broad categories of social actors relevant to the acceptance of the new STS. All actors will exhibit different degrees of support or opposition toward any of the socioeconomic and technologic aspects of the STS of energy provision that includes wind. For example, the acceptance of wind power by an existing power distributor may be highly determined by the value it places on the variability of wind power. For such a company, acceptance concerns the specific wind-related pattern of energy production, which they perceive as complicating reliable provision. Their perception of variability is different, and discussing this attribute they are “framing” it differently: they tend to define this variability as “intermittency” [27]. Their acceptance also concerns completely different aspects, such as their attitude toward newly emerging energy producers like farmers, corporations of civilians investing in wind power, or newly emerging companies that focus on the development of renewables’ installations. During the discussion of the three dimensions of social acceptance, the question “who is accepting?” and the actors presented in Table 1 will repeatedly be referred to. In addition to the question “who is accepting,” the other main question “what” is to be accepted will be elaborated.

Subjects of Acceptance

Attitude Objects and Attributes

Acceptance is an act of decision making, based on a trade-off between pros and cons. It is a behavior based on considerations of several aspects regarding the subject. These considerations eventually manifest themselves in a positive or negative disposition (attitude) toward the intention to take action and to say

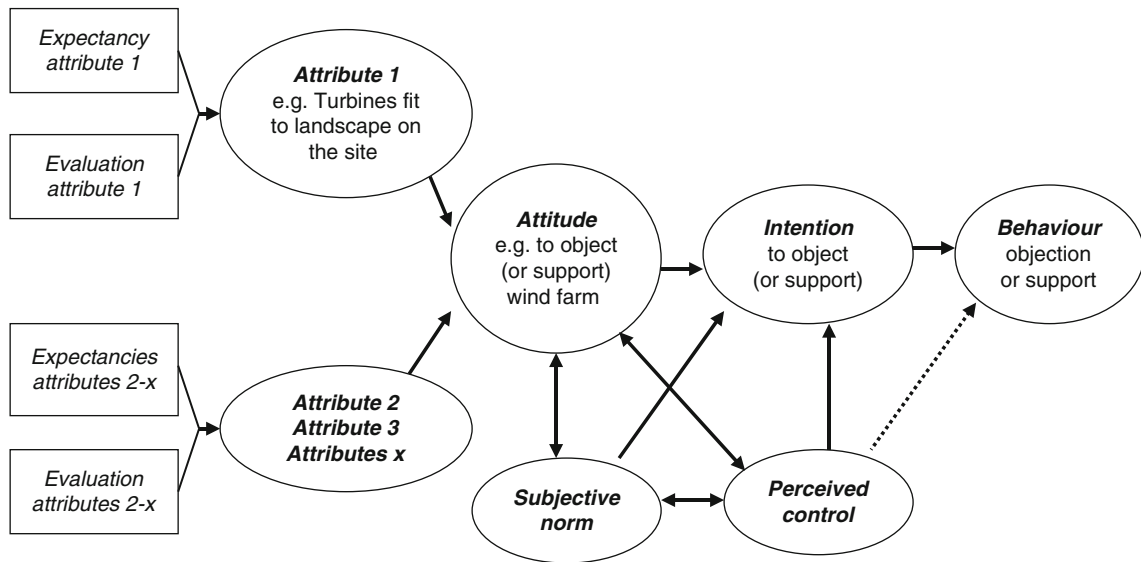
either “yes” or “no” to a certain phenomenon – the attitude object. In psychology, a particularly influential framework on the relation between attitudes and behavior is the Theory of Planned Behavior (TPB) [28, 29]. This model has proven as successful in explaining various types of environmentally relevant behavior, such as choice of travel mode, household-waste composting and recycling, energy saving, water consumption, etc. The model is also applicable to attitudes regarding decisions about implementing renewable energy solutions [30].

Figure 2 illustrates how attitudes and their relation with behavior are constructed according to the TPB. Central is the attitude-object relationship, which in this case is the concrete action of supporting or rejecting a particular decision regarding wind power implementation. Attitudes are constructed according to a model of expectancies and values. The attitude is based on several “beliefs” about the object. Each belief associates the object with a certain attribute; a person’s overall attitude toward the object is determined by the subjective evaluation of those attributes. Although people can form many different beliefs about an object (many potential attributes are presented in Table 2), it is assumed that only beliefs that are readily accessible in memory – “salient beliefs” – influence attitudes. There are many possible actions regarding the application of wind power and all these actions are different attitude objects. From a theoretical starting point, it is fundamental that there can only be clear empirical relations established between attitudes and behavior as both are related to the same object.

The process illustrated in Fig. 1 particularly holds true for attitudes held by individuals. The act of social actors and their underlying positions will generally be based on the same kind of attributes, but the way they are processed in organizations is different from the processing that takes place in the mind of an individual. However, for both types of actors, individual as well as collective, the significance of the existence of attitude-objects relation and several underlying attributes related to that object remains the same. Although it is not possible to simply apply attitude theory and behavior on groups of individuals that together constitute a social actor, the principle remains the same: the actions and positions of social actors also focus on an object, similar to the attitude-object relation for an

Wind Power: Basic Challenge Concerning Social Acceptance. Table 1 Actor categories with relevant acceptance to one or more aspects of wind power implementation

<i>Stakeholders in development</i>	• Incumbents in the existing energy supply sector	– Existing power production companies
		– Power distributing companies
		– Grid managing organizations/ companies
	• Wind power developers	– Existing companies investing in wind as a new activity
		– New emerging organizations or companies, specialized in renewables
	• Wind turbine industry-related actors	
	• Actors with vested interests in domains relevant to establishing wind farms (e.g., R&D, consultancy, engineering, construction, etc.)	
<i>Authorities and public bodies</i>	• Actors representing energy consumers' interests	
	• All actors with secondary interest in investments in wind power (e.g., financial)	
	• National government	– Ministries in policy domains relevant to wind power implementation
		– Energy market regulator(s)
		– Government agencies
	• Regional governments	– Spatial planning officers
		– Regional economic development officers
<i>Stakeholders in related domains</i>	• Local governments	– Spatial planning officers
		– Local economic development officers
	• Landscape-protection organizations	
	• Environment- and nature-protection organizations	
<i>Public</i>	• All actors with interests in competing spatial functions	
	• Actors with interests in economic sectors potentially affected by wind power	
	• General public (electorate, public opinion)	
	• Individuals with any perceived interest in wind developments	
	• Communities (geographically or socially defined)	
	• Civil society organizations representing affected interests	
	• Civil society organizations established because of wind power implementation issues	– Organizations established for private investment in wind developments
		– Organizations established to counteract proposed wind developments



Wind Power: Basic Challenge Concerning Social Acceptance. Figure 2

Formation of attitudes based on expectancies and values about attributes of the object of behavior, in relation to intention and behavior, according to the Theory of Planned Behavior [29, p. 182]. Subjective norm: perceived expectancies of important “others” Perceived control: control of the individual over behavior and its aims

individual. Hence, the question is what are the attitude objects in the case of wind power implementation and which potential attributes exist in those cases?

Two Types of Objects

Within the social acceptance of wind power, many attitude objects exist, because there is a wide range of different kinds of decisions connected to wind power development. For example, an energy company may decide to invest – or not to invest – in a particular wind farm. For consumers, it can also be a decision to invest in a wind farm or to subscribe to a “green power” scheme that includes wind-generated power. For a government actor, it may be a decision to take on a share in the investment in a wind farm [31], to grant a subsidy to investors, to implement a certain financial procurement instrument, etc. Eventually, these decisions seem to converge on the fundamental attitude of the actor toward the implementation of wind power: does an actor have a clear positive, supporting attitude or a negative, opposing attitude toward wind power. However, this is an oversimplification, as for all types of decisions the beliefs and evaluations with regard to wind power are only one of many possible attributes at play.

According to the above-described expectation that there would be strong really existing relations between the attitudes, the intention, and subsequently the decision (Fig. 2), a fundamental requirement is that the objects of attitude and decision fully correspond. However, the attitudes toward a general support of wind power policy and the attitudes toward the construction of a particular wind farm have two completely different attitude objects. As a result, the relationship between the general support for wind power and a decision to support a particular wind farm will be weak. The unfounded expectation that this relation should be strong, reinforced by the commonsense approaches to social acceptance (section “Beyond Common Sense”), has led to strong and persistent misunderstandings about the nature of social acceptance. The two most important misunderstandings concern the fact that public support for wind energy, as expressed in opinion surveys, is high in most countries; however, in many of these countries, the success rates in establishing wind power developments are actually rather low. The two misconceptions are:

- Objections that result from the so-called individual attitude–behaviour “gap”: people are generally in

Wind Power: Basic Challenge Concerning Social Acceptance. Table 2 Attributes of attitudes toward wind power and toward wind power schemes

Attitude object: socio-technical system (power supply with a substantial amount of wind power)		Attitude object: project (wind power scheme)	
Essential characteristics	Associated attributes	Essential characteristics	Associated attributes
Environmentally benign, renewable character	Alternative to fossil	Wind power	Clean character
	Alternative to nuclear		Supply characteristics
	Source cannot run out		Visibility
Supply characteristics	Variability	Location and site	Landscape identity
	Reliability		Nature wildlife
	Capacity credit		Annoyance (noise)
	Domestic source		Wind farm design
			Competing spatial functions
Visibility	Visual: landscape impact of turbines	Investor and/or (co-)ownership	Community in/outsider initiative
	Nature/wildlife: impact on particularly birds		Local economy (benefits/employment)
			Local shareholders
			Community identity
Economics	Price compared to alternatives	Decision-making process	Open/closed
	Economics related to supply characteristics		Participation of public
	Impact on industry and employment		Participation of stakeholders
			Spatial planning procedure Legitimate / "fair"
			Spatial planning procedure
Organizational aspects energy sector	Distributed		
	Decentralized		
	Small scale		
	Options third parties		

favor of wind power, but they change their mind as soon as they are confronted with a proposal;

- The idea that objections to wind developments are simply manifestations of the so-called “not in my backyard” (NIMBY) syndrome.

By dividing the wide range of decisions about wind power implementation into two categories, [Table 2](#) shows the potential objects that are related to

(1) decisions concerning the general application of wind power and the integration of wind-generated energy in our society (on the left), and (2) decisions about establishing concrete wind developments (on the right).

Attitude Object: Energy Source

The idea that wind power as a renewable energy source is a clear climate change mitigation measure seems the

most obvious attribute of the attitude object “wind power.” For this attitude object, the attributes are connected to the characteristics of a new socio-technical system of energy supply that includes substantial portion of wind power. Of course, a prominent characteristic of wind power is its more environmentally benign character compared to alternatives. This attribute is currently dominated by the wish to avoid greenhouse gas emissions, but this immediately shows the dynamic nature of the phenomenon of social acceptance. In the seventies, when several developed countries in Western Europe and some USA states introduced wind power developing policies, climate change was not an argument used to support wind power, neither in public opinion nor among the actors involved in energy supply or policy. There were some environmental arguments, as fossil fuels were clearly associated with air pollution, and also other considerations, as in several countries the public and also some governments were eager to develop alternatives to nuclear power. Clearly, the environmentally benign character of renewable source energy is still a significant attribute for the attitude object “wind power,” but the argument has only recently shifted toward mitigating climate change.

Several studies involving various stakeholders, policy makers, developers, and the public, demonstrated that attributes other than the renewable character of wind are significant for acceptance by all of the relevant actors in these categories [21, 22]. These characteristics and the associated attributes summarized in Table 2 are not only related to the energy source (such as the variability, often framed as “intermittency”) but also to the institutional aspects of implementation of wind energy, such as new options and consequences for incumbents in the energy sector and for national economic and industrial development. Similarly, the price to pay for wind-generated power is obviously important for all types of actors, but for different actors we are dealing with different prices. Their significance is very different in the three dimensions of acceptance (Fig. 1) and they will be discussed in more detail in sections “Sociopolitical Acceptance,” “Community Acceptance,” and “Market Acceptance.”

Attitude Object: Wind Development Project

The object of attitudes toward concrete projects of wind power application concerns two kinds of

decisions: decisions about investing in such development and decisions about building a wind farm at a certain site in a particular community. The character of such decisions is entirely different from general decisions to promote or to restrain further wind power development. As a matter of fact, only one characteristic of these decisions concerns the source of energy. The other characteristics of the decisions are all related to details of the specific proposal. It is important to note that these attributes are only remotely related to the character of the energy source itself, but nevertheless they are at least equally important and often more important than the attributes directly related to wind power as the source of energy.

Some of the attributes in Table 2 of decisions about projects seem to correspond with attributes associated with the object of wind power in general. For example, in both types of decisions, economic arguments play a significant role. However, the economic arguments for applying wind power and its impact on general welfare and industrial development are very different from the economic valuation by potential investors in a wind power project, or from the local community’s beliefs about the existence and valuation of the project’s contribution to the local economy. Similarly, at the general level, landscape issues are certainly very important for determining the attitudes toward wind power development. Here it concerns a general belief that large-scale wind power implementation will result in a myriad of wind turbines popping up in the landscape and the valuation of this phenomenon. Usually, this aspect is neutrally referred to in terms like “visual impact” [32], but in most studies any impact in terms of visibility is generalized and considered beforehand as “intrusive” or “disturbing,” as if this valuation applies to all types of landscapes and also to all individuals. The landscape evaluation discussed in section “Landscape/Seascape” shows that the visual impact of wind turbines varies significantly among both individuals as well as cases, and that it is not only evaluated negatively. In decisions about a particular project, the general assessment is translated into “landscape identity,” which is something very different as it is not primarily determined by the common denominator “wind turbines” but rather by the landscape characteristics of the site of that particular project as they are perceived by the people that feel attached to that site. As these

landscapes vary widely among different project cases, they cannot be considered as principally determined by the general landscape assessment of wind power.

A similar “translation” takes place in the organization of energy supply. For wind power in general, as is discussed in section “[Sociopolitical Acceptance](#)” on sociopolitical acceptance, the option for third parties and new actors to invest in wind power is a relevant attribute ([Table 2](#)). In the translation into a particular project, this becomes a question of wind farm ownership and community involvement, or the acceptance of investments and ownership of a wind power development by companies from outside the community [33].

The large differences between the relevant actors and attributes in the decisions about applying the energy source and the decisions about building wind farms result in only a weak correlation between support for wind power and support for wind farm proposals. The limits of this correlation may be the origin of what is usually referred to as the “gap” between general high acceptance of wind power and the perception that a large number of decisions about building wind farms are failing [34]. The latter perception is problematic, as it is hard to define a basis to establish number of failing and successful projects. It fully depends on when an idea to build a wind farm is considered as a “project” that can either become a success or a failure. Official numbers can be derived from planning consent and building permits, but the success rate becomes high for countries with a planning system in which permit applications start after – unofficial – negotiations, whereas in other countries an application is the start of the process. Example figures from the UK during a 5 year period showed that 60% of applications were eventually refused, while 40% were approved [35].

Again, such figures are highly depending on the context of the country and its planning system and practice, but at least they show that it is far from self-evident that a wind scheme will be realized. However, it is questionable whether this phenomenon indicates that “gap” between general high acceptance of wind power and refusal to support local developments really exists. In fact, the limited correlation between support for wind power and support for wind farm proposals is poorly recognized and understood by wind power advocates and policy makers alike. The fundamental

question regarding this “gap” is why decisions about a particular project, which are taken by local actors and the interested investors, would be highly affected by the fact that national public opinion shows high positive responses to wind power. A clear correlation between general national public opinion and local decision making is theoretically not supported. Even within a single actor, there is a distinction between the attitude with an attitude object (wind energy) that is entirely different from the object of behavior (taking decisions about investing in a wind farm and siting it). [Table 2](#) shows that both objects have many different attributes; therefore, the fundamental condition that the objects of attitude and behavior must be the same to find a clear relation between attitude and behavior ([28]; [Fig. 2](#)) is not fulfilled. The commonsense idea that a decision regarding building a wind farm is primarily informed by general attitudes on wind power, is unfounded by what is known about the relationship between attitudes and decision behavior. Speaking in terms of attitude-behavior theory, the objects of attitude and behavior simply do not correspond.

Mind the “Gap”

With regards the assumed “gap” between general acceptance of wind power and still a low number of positive decisions to build wind farms, Bell et al. [34] distinguish between a possible “social gap” and an “individual gap.” The “social gap” refers to the high-level public support for wind energy expressed in opinion surveys and the low success rate achieved in planning applications for wind power developments.

The individual “gap” emerges from the idea that many of the failures to establish wind developments arise from the fact that individuals, in particular local residents, have a positive attitude to wind power in general, but actively oppose a particular wind power development when confronted with a proposal. The most popular commonsense explanation for this assumption is the claim that NIMBY-ism is responsible for this gap, but Bell et al. already explain the weaknesses of the NIMBY claim in explaining an attitude-behavior gap [34, p. 465]. This popular label is briefly discussed in section “[The Backyard](#),” as it is currently still prominent in commonsense thinking about acceptance of wind power.

However, the terminology of an existing “gap” may also be considered disputable. Table 2 shows the different potential attributes that can be connected to those objects. The comparison of these two categories already reveals a large divergence between the objects in the two columns. The attitude object “wind power” (left) is only one of four main characteristics of the object “wind power project” (right). Attitudes toward a concrete proposal for the construction of a wind farm at a specific location and toward investing in a wind farm include many potential attributes all related to variables that show a wide variety among cases. The fact that this proposal is an attempt to implement wind power is only one of these many attributes; some of the other attributes are often more important for key actors than the implementation of wind power. Examples of potentially more important attributes include the type of the developer – from a large energy company to a nonprofit community associated initiative –, the landscape of the chosen location, the connection of the wind farm to the local economy, the potential environmental impact, etc.

From a developers’ point of view, the struggle to site a wind power development is easily interpreted as a problem of local inhabitants trying to safeguard their environs against turbines [2, 36]. The number of documents reproducing the common perception that any resistance these developers face is an indication of a so-called NIMBY-syndrome is overwhelming and this commonsense view is continuously reproducing itself. It is prominent among the strong proponents of wind power and presented as a “fact” [36, p. 299]. As this idea is engrained, some discussion is required, above all because it is a perfect example of institutionalized thinking patterns that eventually become barriers to the acceptance of wind development schemes.

The Backyard

Many proposed wind developments indeed do encounter resistance, but this cannot be considered evidence for the existence of a NIMBY syndrome. Stronger evidence is needed that the individual’s motivation to resist is as selfish as the use of the term NIMBY suggests [36]. The problem is that such evidence has seldom been found, but still the interpretation of any opposition in terms of the backyard is repeated over and over

again [37, 38]. Any reference to an alleged “not-in-my-back-yard” syndrome suggests a degree of selfishness, conservatism, unwillingness to cooperate, or ignorance of environmental issues. For example, with regard to renewable energy facilities, Kahn [39, p. 26] gives a fairly clear characterization. A NIMBY opponent would be

- *Democratic* with a small “d,” since his/her political right of self-determination is exercised to the fullest
- *Parochial* because he/she is unconcerned with broad impacts
- *Reactive* in that he/she is not generally concerned with planning or siting issues per se
- The NIMBY – opponent is *not an environmentalist*

Although this view is still very prominent among developers and policy makers, with the start of public and social acceptance studies in the 1980s, researchers started to question the relevance of this rather normative NIMBY label. The first notion to be reluctant with the use of this label suggested that it is dangerous for authorities or utilities to use this acronym. It tends to offend the public and will generate stronger opposition and the use of the qualification may easily become counterproductive [40]. The indolent use of the NIMBY label has become a tool to question the legitimacy of the opposition and the actors who represent it. This is a violation of basic assumption IV about the legitimacy of all views on wind power implementation. Meanwhile all efforts of authorities and developers to narrow local concerns about schemes that often have been developed without involving them have become an indication of sociopolitical nonacceptance. If policies really want positive local decisions on wind power schemes, the acceptance in implementation in policy of assumption IV is crucial.

The last two decades, the “theory” of the backyard-motivated resistance to wind developments has increasingly become discredited by researchers: “The NIMBY concept that has been most frequently used to describe negative perceptions of wind farms has failed to receive empirical support” [38, p. 136]. In fact, all researchers investigating the motives for resistance to wind developments, in light of the general public support for the technology, reached a consensus on NIMBY [34, 38, 41–43], which is succinctly summarized as follows: use of this term does not explain the opposition. There are three general reasons against the use of

this term. “First, it is generally used as a pejorative implying selfishness as an underlying cause; second, it appears to incorrectly describe much of the local opposition to wind projects; and third, the actual causes of opposition are obscured, not explained, by the label” [44, p. 124]. Currently the most comprehensive overview of research of community acceptance of on- and offshore wind power — also including other cases of energy innovation infrastructure such as tidal, wave, marine, hydrogen, and bioenergy — discards it as follows: “Fundamentally, the backyard concept is built upon an impoverished view of the ‘home’ (. . . .), local concerns about the impact of renewable energy projects clearly transcending the boundaries of specific properties, or implicating wider spatial domains such as publicly owned landscapes and seascapes” [45, p. 62]. The alternative put forward in that study to think beyond the backyard is place identity, a concept that emphasises the enormous variety of geographical identities of places that can be used for wind power development, which includes landscape as well as community identity variables (section “[Identity of Place](#)”).

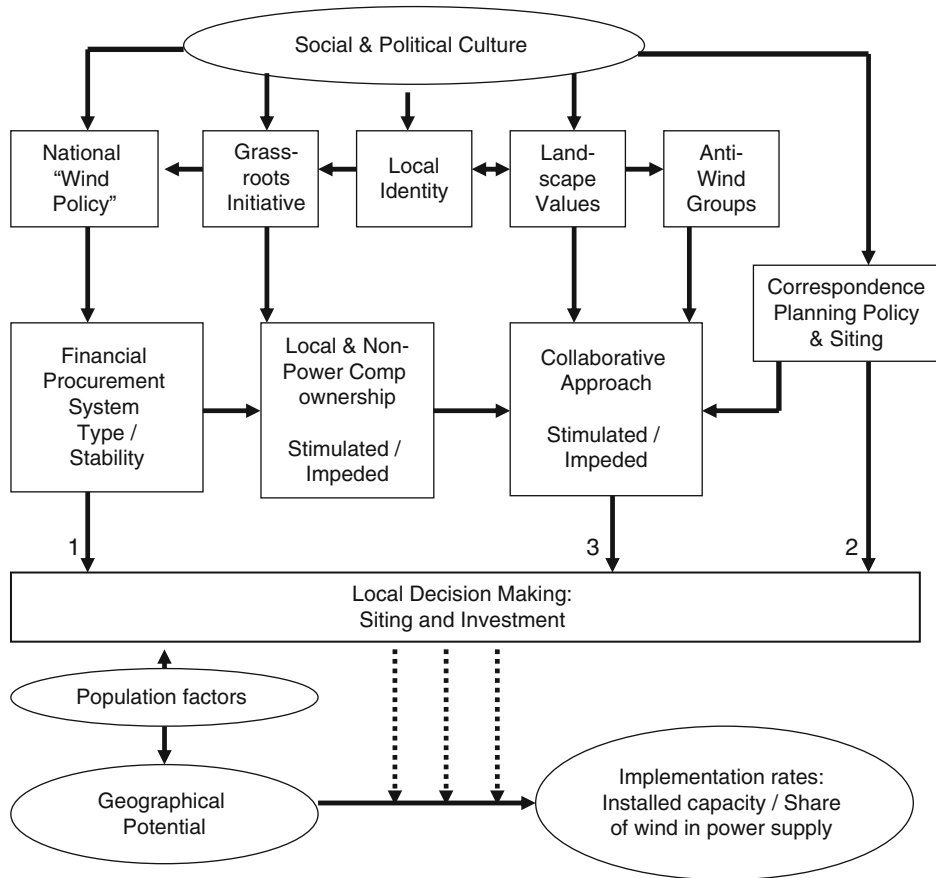
The erosion of the assumed relevance of NIMBY-ism is not only due to studies on wind power. For example, a study on the acceptance of nuclear waste storage facilities also questioned its applicability, as it observed that the users of the NIMBY label do not apply a clear definition and hence, do not make clear what they really mean when they use the label. A clear definition is needed, because if all opposition to facilities were NIMBY-ism, then the concept would be both quite empty and unnecessary [46]. However, when it is clearly defined what motives are distinctive for NIMBY-ism, there is hardly anyone found to whom the definition applies [46, 47]. The most distinctive definition of NIMBY is “an attitude ascribed to persons who object to the siting of something they regard as detrimental or hazardous in their own neighbourhood, while by implication raising no such objections to similar developments elsewhere” [48]. The second part of this definition is crucial, because it distinguishes NIMBY from opposition motivated by other factors as it explicitly relates the opposition to merely the fact that it is situated in the backyard. Most of the manifold critical stands are related to this aspect, because that people support similar development elsewhere is not supported by empirical evidence. The assumption

implies that close means opposition, far away is support. The most comprehensive effort to formulate a theory of NIMBY postulates as the most fundamental hypothesis that the universal factor is geographical proximity. “The rule is simple: The closer residents are to an unwanted facility, the more likely they are to oppose it.” [49, p. 290]. This idea is circulating in policies and publications so frequently, that some researchers seem really surprised to find exactly the opposite. The general truth is that, in some cases, this relation between opposition and distance is actually found — though even these studies emphasise that this should not be considered a result validating the backyard these [50] — but more often researchers find that people close to the wind developments are on average more in favor of them (e.g., [69, 93]), and the proximity hypothesis must be considered falsified.

Developers and policy makers, in particular those in countries that have not been successful in establishing institutional frames that facilitate rapid implementation rates, are still inclined to think in terms of “backyard” attitudes of the local public [21, 39]. The persistence of this inclination is probably due to the fact that ultimately the local level is the level where the real decisions about investments and building concrete wind power farms are taken. The institutional factors that affect decision making on that level are summarized in [Fig. 3](#) [51]. It is good to note that factors such as “population density” and “geographical potential,” including the geographical distribution of wind resources, are clearly informing the possibilities for establishing wind farms. As described in the section “[Institutional Change](#),” the focus here is on the institutional factors that are found in the upper part of [Fig. 3](#). The key question is how decision making on the local level is organized and how social networks, operating on either the same or other levels, are influencing those decisions. Market acceptance and community acceptance are mainly connected to this level, whereas essential factors that are framing those decisions are located at the level of general sociopolitical acceptance.

Sociopolitical Acceptance

The three keys elements to securing to positive decisions about establishing wind farms are:



Wind Power: Basic Challenge Concerning Social Acceptance. Figure 3

Overview of institutional factors affecting the decisions about wind power implementation (Adapted from [51, 52])

1. The shape and the reliability of the financial procurement system (Fig. 3: 1), which is of overwhelming importance for the utilization of the potentially significant willingness to invest in wind developments;
2. The type and amount of effective support generated within the national and regional planning system for policies that develop wind power capacity (Fig. 3: 2);
3. The degree to which collaborative decision making on the level of communities is allowed and stimulated by the sociopolitical framework (Fig. 3: 3).

The third factor is closest to the central question in sociopolitical acceptance, as this factor is partly determined by the two others, the financial procurement system and the planning regime. Collaborative planning and decision making [53] is the favorable

condition for reaching positive decisions about wind farm construction, but the conditions that determine whether or not collaborative approaches prevail at the community level largely depend on sociopolitical acceptance. The willingness to establish effective support regimes in these domains varies greatly. Clearly visible in Europe (Fig. 3 is the result of an analysis of six EU cases) [51, 54], it has also been equally significant in the comparison of the effectiveness of policy frameworks among states in the US [55, 56].

There are four institutional variables which concern decision making above the community level: the institutional and financial support system; the planning regime that should be in line with the support system [57]; the value attached to landscape preservation; and the framework determining the degree of local ownership of wind power schemes. Of these four,

landscape values are connected to history and largely culturally determined, and exercise great influence on decision making regarding wind power schemes (see section “[Landscape/Seascape](#)”). It is one of the reasons why a collaborative approach in decision making is so important. The other three factors are mainly informed by the institutional developments in the dimension of sociopolitical acceptance. These three constitute the institutional capacity that can foster and exploit the high average support for wind as a power source ([Table 2](#), left). Some countries have shown limited learning to develop this institutional capacity [58]. Most still rely upon the assumed high level of acceptance among the public. This support is very strong in most countries, but it is important to notice that in all countries there is a minority that does not support wind power and, furthermore, support is almost always conditional [34]. In order to meet the conditions for that support, serious institutional change is needed, but those changes are subject to sociopolitical acceptance.

Institutional Conditions

The main problem with sociopolitical acceptance is that it is hard for policy makers to recognise basic assumption II of social acceptance, the institutional nature of WP implementation, especially the sociocultural side of the STS. Within policy, there is a strong rooted institutional tendency to look at energy issues and at environmental issues in general primarily – may be even solely – as technical problems. Also experts see the “energy problem” as a chiefly technical issue, and policymakers tend to continue to overly rely on technology for providing solutions to issues that it cannot address, namely, deeper social, political, and cultural problems [59, 60]. If one assumes that policymakers are willing to promote wind power, this inappropriate focus remains an obstacle to looking at the technologies to be socio-technical and consequently policymakers continue to promote renewables ineffectively [20]. Although many developed countries have been promoting wind power for more than two decades and also have instituted policies in support of wind power, there are enormous differences in the effectiveness of these policies.

The comparative assessments of countries and state-specific sociopolitical contexts provide some

understanding of the policies and processes that affect wind deployment. For example, a comparison of four US states established that although the state of Massachusetts implements supportive policies, this support has been insufficient to trigger deployment as deployment costs remained very high. On the other hand, Texas’ targeted policy and spatial planning fostered policy that was in line with the energy policy objectives, and facilitated deployment. Furthermore, the positive aspects of community-based wind development have been shown, also with regard public acceptance, in particular in Minnesota [60] and Texas [50].

Related to the factors in [Fig. 3](#) (discussed in the next sections in greater detail), it is important to emphasize that the existing discourses on all aspects of sociopolitical acceptance among all important key stakeholders show fundamental differences, for example in Europe [21], the UK, and Australia [22]. Also in the US, the nature of wind energy discourses differs significantly within states [60]. These contrasting discourses reveal strong deviating beliefs about the necessity of wind-generated power as well as fundamentally deviating beliefs about how to implement wind power. These differences reflect separately operating and usually competing advocacy coalitions within the domain of policies that are relevant to wind power. Within the conflicting policy belief systems of these advocacy coalitions [61], the problematic nature of sociopolitical acceptance is reflected [21]. Within different discourses, actors usually also use different terminology for basically the same attributes, phenomena, and issues. All actors have vested interests, and almost any actor has inherent tendencies to use specific terminology to create a certain perspective that fits to their views. This process is called “framing” and this has proven to be effective, because individuals perceive losses and gains differently, depending on how a phenomenon or issues is presented to them [62]. Some examples: some may define financial procurement as “financial support,” whereas others prefer to frame it as “subsidies”; the fact that wind power supply is not following demand but the presence of wind may be called “variability,” but it is also presented as “intermittency” [27]; the continuous power supply of large-scale power from nuclear may be presented as “reliable,” but it can also be framed as “inflexible.”

The strong differences in policy beliefs reflect the level of acceptance for the change of essential institutions, for example the willingness to support the choice for and the implementation of effective regulatory regimes and the willingness to establish planning regimes that create the necessary environment for collaborative decision making at the local level. Besides overcoming the resistance from strong stakeholders, in particular in the energy sector, it is not easy to introduce supportive regulatory systems. Supportive regulation is complex: changing a single element of a current system, for example only the financial support system, or only the planning regime, will not overcome barriers at deeper political, cultural, and social levels.

Public Opinion

What is usually seen as most problematic, general public acceptance of the energy source, is in fact least problematic [23]. Several indicators show, from the beginning in the 1980s and 1990s onward [63] that public acceptance of renewable energy technologies and policies is high in many countries. General public opinion polls in Europe show that the majority agrees with public support for renewables, also in countries where the government provides relatively little support [64]. In the US, a majority also supports building wind facilities. In a study on different power-generating facilities, it was found that attitudes about plant siting depend heavily on perceptions of environmental harm and costs of the particular facilities, but the effects of these attributes are similar across different types of fuel sources [65]. This suggests that people view all power sources in the same framework and differentiate them on perceived endowments, environmental harm being the most important. Thus, both solar and wind power enjoy high public rating compared to other sources [64, p. 27]. This positive overall picture of public support for the wind power as energy source for a long time misled policy makers to believe that social acceptance is not an issue. This is a mistake, for three reasons:

1. General public support is only based on attributes with regard to the energy source in general and not connected to attributes that are cases specific (Table 2).

2. Public support is based on fairly low information levels [66]. As a result, when the public becomes better informed, their views become more balanced, for example when there is a wind power proposal and a public discussion about it within their community [67]. In fact, because of the very high level of positive attitude in the poorly informed initial condition, the average shift as a result of becoming informed shows the pattern of a regression to the mean (see the end of this section).
3. Public support may be averagely high, but this average still contains a minority of people that generally do object to wind energy. In all studies on discourses, the objectors to wind energy appear as a distinguished class of actors, not only among the public but also among the key actors in wind power decision making at all levels [21, 68]. In concrete cases of wind farms proposals, the part of the public and the actors that represent this objectors' discourse will almost certainly participate. Because of the strong emotional component in landscape valuation, their involvement in decision making is usually very intense, because landscape is the main attribute for attitudes toward wind power projects (see section "[Landscape/Seascape](#)").

The commonsense focus on opposition as a barrier to wind power development implies a "taken for granted" view on the willingness among the public to take positive decisions about wind power. While the willingness to invest (positive investment decisions) and the willingness to integrate wind power applications in the community (positive construction and application decisions) are conceptually very different. In practice, these two are sometimes treated in a similar way within existing institutional arrangements. For example, willingness to pay for "green power" [69, 70] is by no means similar to willingness to invest in a wind farm and even less similar to a positive attitude toward a particular wind power scheme. Therefore, a positive attitude toward "green power" or "green tariffs" is not a proxy for public attitudes toward renewables in general and even less for attitudes regarding wind power (see section "[Market Acceptance](#)").

A particular feature of public acceptance is its time dimension. The typical pattern of acceptance before,

during, and after a project follows a U-curve, falling from high average general acceptance to more critical levels of acceptance during the siting phase (usually still positive on average) and back up to a higher level of acceptance once a project is up and running [41, 67]. Remarkably, this effect is due to the low levels of public knowledge in the first stage, when people are not yet confronted with any wind development in their environs. Information and discussion about a concrete project in the community is increasing knowledge. This is making attitudes more nuanced and therefore averagely less extremely positive in cases where no high emotional attributes prevail. In other cases, it may also result in more polarization, which also means a lower average positive attitude, when strong emotional attributes – community and landscape identification related – are at stake [68]. Again, this attenuation effect on the generally positive attitudes as a result of increased information levels runs contrary to common sense, as developers and authorities often expect that increasing knowledge will help to reduce objections against wind power projects [71]. However, this expectation is based on an illusion, as there is no study that provides any evidence of a positive relationship between the level of knowledge about wind energy and attitudes toward wind energy or toward a particular wind power project.

Whereas the general public support does not imply strong impediments for wind power implementation, the following set of institutional domains in policy development and energy provision are the distinguishing determinants for the level of sociopolitical acceptance.

Economic Conditions and Financial Support

Support regimes and their stability provide the basic premise for successful wind power programs, and their perceived stability has a positive effect on investment decisions. The support regimes are primarily financial, and obviously the substantial incentives provide a stronger support than small incentives. However, the institutional conditions of the support systems are even more important: do they subsidize capacity or generated energy?; do they support existing incumbents or do they attract new market actors?; are the subsidies ultimately paid by tax payers or are they

guaranteed rates ultimately paid by energy consumers? These essential characteristics affect the effectiveness of the support as well as all three sorts of social acceptance.

In most countries, the dominant approach has been to stimulate investments with the introduction of government subsidy or by deploying a tax credit. However, the renewable energy feed-in tariff (REFIT) system, as a financial procurement system, has become a cornerstone for the deployment of wind power in successful countries [72, 73]. The effectiveness is due to the fact that this system is not only favoring investments – market acceptance – in wind schemes, but it is favoring investments by new actors in energy production, which is also supportive to collaborative planning of wind power schemes because those actors often have strong connections or are rooted in the community (Fig. 3).

The clear example of the promotional role of this key instrument in fostering social acceptance is the innovation of wind and solar technologies in Germany, which is linked to the rapid diffusion that resulted from the nature of the policy instruments employed. However, the application of a clearly effective policy instrument like REFIT is everything but easy; most countries did not implement it and in fact its implementation is subject to sociopolitical acceptance. Analysis of the sociopolitical process that led to the adoption of these instruments in Germany demonstrates how the regulatory framework is formed in a “battle over institutions”: “The German parliament, informed and supported by an advocacy coalition of growing strength, backed support policies for renewables-sourced electricity against often-reluctant governments and the opposition from nuclear and coal interests” [74, p. 256]. On the one hand, this example illustrates high acceptance as the instruments that were implemented were very effective indeed, but, on the other hand, the example also reveals that there are strong forces that provide sound resistance in the policy process. Such resisting stakeholders exist in all four categories of actors (Table 1), but with regards the financial procurement system and the guaranteed access to the grid, the strongest resistance is often found among the incumbents in the energy sector itself [51, 75]. This feature of access of new market actors in fact reflects the choice to prioritize power generated

with renewable sources over conventional power generation, and the appearance of new actors on the market is a key in innovation processes [8, 12]. The incumbents in the power sector prefer a system based on Renewable Portfolio Standards (RPS) with a claim that such systems are market based. Although they do not create an electricity market (unlike REFIT, which creates this market based on fixed prices) and only offer certificates, RPS supporters claim that it is more market based and more cost effective [74, 76]. However, several studies have shown that REFIT is at least far more effective, and some also claim more cost effective [77, 78]. The RPS systems actually limit the market to a few large energy companies [75, 79].

National policy defines the objectives, conditions, and instruments of the financial support systems. Most significant is the fact that REFIT grants grid access to all and it provides financial risk reduction for new investors [76], and therefore it supports new, more inventive, and renewable-friendly players, thus enhancing market competition. Furthermore, the emergence of such market competition puts pressure on the incumbents. This is a crucial positive reinforcement for any civil society initiative that seeks to promote renewable energy. In order to exploit the large potential of supportive attitudes in the society, favorable conditions and institutional change are needed. The introduction of REFIT systems was a major effective institutional change in some countries, but many countries still refrain from introducing such systems. Moreover, despite its undeniable effectiveness, the REFIT is continuously under attack from energy companies and several EU member state governments. The new regulatory regime in Germany has gained widespread support, but resistance still exists and has sought to influence European regulation in order to replace the regulatory regime adopted in Germany [79, 80].

The most significant institutional changes – changes of the “rules of the game” – are achieved by political decisions, so the roots of the huge differences in success rates of wind power implementation can be found in primarily problematic sociopolitical acceptance. In some countries, sociopolitical acceptance of effective new regulatory regimes has been high and accompanied by policy changes, in many other countries it has enjoyed only rhetorical support. Expressed preference is not always evidently followed by action

among policymakers and structural measures – institutional change – are yet to materialize.

The Impact of the Spatial Planning and of Location Decision Making

A persistent line of explanation for the implementation deficit that several countries experience with regards their performance achieving substantial wind power capacity is the existence of a “planning problem” [71]. This implies that spatial planning would constitute a “barrier” to the expansion of wind energy. Most of the attention in social acceptance studies has been focused on studying the continued deficit in implementation as a result of problematic local decision making [8]. From the developer’s perspective, the planning problem is usually interpreted as an attitude-behavior “gap” between general public support for wind power and local opposition to wind power schemes (section “Mind the ‘Gap’”). The reaction to these problems is often a request for more perseverant planning instruments that would enable authorities to overrule local opposition. However, the nature of the planning problem is strongly contingent on the discourse that the actor supports, and the most prominent followers of the “planning as a barrier” line are found among developers and authorities in countries that have not been very successful in implementation [21]. Actors in more successful countries have learned that collaborative planning approaches that foster civil society initiatives and involve local stakeholders generally produce an increase in positive decisions about investing and siting wind farms.

The request for a more powerful, in fact top-down planning regime produces a uniform, inflexible system. Motivations for adopting such approaches may also reflect centralizing planning cultures. A clear example of this is the Welsh approach to the planning problem, as a reflection of the UK’s centralizing tendency in planning. The Welsh 2005 planning guidance on renewable energy superimposed centrally determined “Strategic Search Areas” for large-scale, onshore wind farm development onto local decision-making processes [81]. It is doubtful whether such increased centralization will help accelerate wind power implementation. In a comparison of UK and Australian policies, it was concluded that such policies did not

ease conflict over wind energy projects. On the contrary, they have been criticized as endorsing wind energy while neglecting concerns about landscape and place [22]. Another example is Sweden, where the main purpose of the national planning instruments is to strengthen wind power as a public interest in local land-use planning processes. According to Bergek [82], this purpose is not well served because civil servants became more inclined to treat wind power as a private interest when the country targets are achieved.

If national planning uniformly superimposes the locations and site selection, it can only apply the general attributes of wind energy (Table 2; left column). In decision making regarding projects, these constitute only a part of the attributes (Table 2; right column) that are significant to the actors involved in the process. In the practice of such decision making, another UK study showed how the attitudes of people in the immediate vicinity of proposed wind farms exercised the most important influence on the decisions made by local authorities [83]. This not only includes the perception of the appropriate match between the wind farm and the character of the landscape (section “[Landscape/Seascape](#)”) but also concerns the perceived local economic benefits. Interestingly, the study also concluded that there is a lot that wind power developers can do to improve the prospects of successful planning. This is where the endorsement of national policies through top-down planning may easily become a factor creating mistrust (see section “[Community Acceptance](#)”). In addition to landscape concerns, negative attitudes also emerge in response to the dissatisfaction with the planning and decision-making processes [84]. The opposition to wind power developments may even be shaped largely by the manner in which planning and decision making on wind schemes takes place [67, 85], for example the manner of organizing public consultations [86]. The issue is that all local cases of wind power implementation show a wide variety of place identification factors, because both the landscapes as well as the social structure of the host communities differ greatly. Hence, the attributes of wind power developments as listed in Table 1 will be assessed differently in all cases. Uniform policies cannot address this diversity. Studies of the underlying reasons for opposition show that the objections are very heterogeneous [68]. As wind power is already

a contested realm, also the legitimacy of national policy is problematic, in particular when hierarchical power is used overtly to site turbines.

In case planning is actually organized with a focus on local decisions and not primarily based on national uniform policy regulation, the consequence is that the planning context may vary between different municipalities, as community acceptance levels may be different and may also vary in time. This poses a particular difficulty for larger, nationally or even internationally operating companies. It is also not congruent with national uniform systems that claim to support wind power in a cost-effective way. As was noted in an analysis of Swedish wind power planning, municipal planning tends to undermine the cost-efficiency goal of the “green certificates” trading systems [87]. In Germany, the feed-in support system – REFIT – has positively affected the involvement of the civil society and local initiatives to wind farms and correspondingly. The implemented planning system did not endorse the selection of sites by higher authorities; instead it only set a priority for wind developments (“*Privilgierung*”) and then left the actual planning of the local wind development and site selection to local governance [88].

There are huge differences in national planning systems. In general, it can be concluded that planning systems are not creating the problems of social acceptance. Planning is not the source of the problems, but in many cases may become the carrier as the planning process is framing how local decisions are taken, and most issues related to acceptance manifest themselves in the planning process. Therefore, supportive planning systems are generally those that effectively foster and enhance market and community acceptance, i.e., a system that stimulates collaborative decision making at the level where crucial decisions about establishing wind farms are taken. Substantial evidence points to the many benefits of the early, sustained, and reciprocal engagement of host communities. Such strategies have been linked with improved chances of planning success and are seen as offering developers the opportunity to gain the trust of host communities and among actors that invest in wind power projects, to identify and address their concerns and to effectively communicate the potential risks and benefits. This does not happen on the national level but rather mostly on the local, with the notable exception of Spain [31].

Community Acceptance

Two Dimensions of Community Involvement

Historically, the focus within this issue of social acceptance of renewable energy innovation has been on public acceptance as the cornerstone of community acceptance. Community acceptance refers to the specific acceptance of siting decisions and renewable energy projects by local stakeholders, particularly residents and local authorities. In the commonsense approach of “nontechnical factors” (see section “[Beyond Common Sense](#)”), it was soon recognized that energy companies trying to develop pilot projects faced problems with the communities. Because they were usually incapable to handle these problems, about the local resistance in the communities where they wanted to establish those developments, the debate around NIMBY-ism unfolded. Section “[The Backyard](#)” describes that this classification of local nonacceptance is not fruitful as it violates Basic Assumption IV by questioning the legitimacy of the actors’ positions and because it fully neglects the high potential of positive acceptance at the community level of initiatives that originate within the local community itself and in civil society more broadly. The utilization of such labels denigrates opposition to proposed wind development as cynical and selfish.

Most case studies that analyze the motives and beliefs of local actors reveal a wide range of considerations, ranging from strong support for local initiatives to rejection of wind developments in principle. The essence of the difference between the positions that people take regarding specific wind development projects and the support for application of wind power in principle is that the latter general support is qualified [23, 34, p. 463]. All actors in society may be prepared to support any wind development that meets certain criteria, and they will tend to oppose developments that do not meet those criteria. From the developers’ point of view, their main concern is the construction of the wind turbines, but other actors have other concerns: the selection of a site out of many alternatives, different options for who is the investor and owner of the turbines, and considerations on who will decide about the project and how. For the developer, these aspects are not variables but self-evident constants. They are inclined to present the proposal in a way

that creates a narrow frame for the community. The developer is not inclined to look at himself as a variable, as just one out of many alternatives. Usually, the proposal is a wind farm at a chosen location with an already selected site. For the community, this manner of planning a wind power scheme (“decide – announce – defend”) is seen as “framing” the decision-making process: it presents the most significant aspects of the project as already decided upon, without their participation [67].

This “framing” of decisions of wind power is rooted in several institutions, such as rules and procedures in spatial planning systems, and existing preferences in energy policies that favor developments initiated by the incumbents in the energy sector. The process of altering these types of institutions is part of sociopolitical acceptance (section “[Sociopolitical Acceptance](#)”). Moreover, these kinds of problems do not remain restricted to wind power. Local acceptance of all types of infrastructure facilities is steeped in local conflict [89] and infrastructures that aim to further environmental policy targets, such as environmentally benign waste management, and climate change adaptation or mitigation are no exceptions [90]. The roots of these conflicts are complex as they always combine opposing interests regarding the use of the facility and the burden of the facility in terms of local social or environmental impact. Considering wind power, four different types of conflicts of interest may be recognized, which are based on two dimensions:

1. Application of wind power as a public interest versus wind power as a private interest of the owner of the wind farm
2. Wind power in conflict with various private interests versus wind power in conflict with public interests of the community [82, p. 2360]

Actors of all four categories listed in [Table 1](#) are attached to any of the four types of conflict. By definition, developers represent private interests, except when local or regional authorities participate in wind farms, as occurs in Spain [31]. Nevertheless, developers often claim that their project is an application of an environmentally benign energy source and represents a public interest: “The benefits of wind power are felt by the masses in improved environmental quality, diversity of energy resources, and compliance of

international agreements [..but] the costs are borne locally.” [91, p. 505]. In case the developer is a community outsider, such as an incumbent energy company, the local perception will assign the developer as representing private interest, whereas the impact is perceived as a burden to the community’s public interest [33]. When developers emerge from within the community, their interest may be private (e.g., a farmer) but it may be public when the developer is a cooperative rooted in the community, as frequently has been the practice in success stories like Denmark and Germany. In addition to these interests, there is the second dimension of community involvement: the decision-making process that may vary from closed and institutionally constrained to open and participatory. On both dimensions of “community renewable energy,” the utility wind farm scores low, with severe consequences for its acceptability from the community’s viewpoint [33, p. 498].

An illustration is given in a recent study in Germany, showing clear differences in acceptance between two fairly similar communities, including the visual impact and concern over environmental protection and high energy use. A community energy initiative in a village called Zschadraß, changed the perception of the local people toward renewable energy compared to a control community – village called Nossen – without such an initiative. The residents of Zschadraß had become consistently more positive about renewable energy: “Community co-ownership of wind energy leads in fact to a higher level of local acceptance toward the nearby installed wind turbines than ownership of a wind farm by a commercial company” [92, p. 7]. Some significant figures: the statement “There are already enough wind mills in the area” was supported by 39% in Nossen, and only 8% in Zschadraß, and comparable figures on the statement “...and that causes a disturbance of the landscape” were 18% and 4% respectively.

Identity of Place

Related to the local acceptance of wind power projects is the question whether institutional settings foster the community’s identification with the project, such as locally organized or publicly owned wind power. This can obviously be the result of community involvement in the outcome of the project, which can take the form

of ownership or shareholder participation. However, the identification can also be more symbolic “sense of ownership” [93], as a result of effective involvement and influence in the siting process.

Moreover, all the forms of involvement are part of or are supporting collaborative decision making at the level of wind power scheme (Fig. 3: 3). How decision making is organized and how social networks at this level are involved in projects strongly shapes the possibilities for identification with the project by all actors at the community level [94].

The most frequently mentioned commonsense interpretation of identification is the proximity of the selected wind farm site; this proximity hypothesis is closely associated with the “backyard” explanation of rejection of wind farm development, i.e., local attitudes to wind power would be negatively correlated with proximity to wind turbines. However, studies have failed to find evidence for the presumed negative effect of spatial proximity on public attitudes. Although some studies did indeed find more negative attitudes at closer distance [50], several studies found – sometimes to the researcher’s surprise – that attitudes to wind power as well as to the existing project became more positive at shorter distances [69, 93]. These results also seem to be independent of the national policy context. For example, a recent study in New Zealand found similar results to the previous studies in Europe: neither was there a clear relationship between the proximity of the submitters of appeals to the proposed wind farm and their appeals, nor did proximity increase the likelihood of them having negative perception of the proposal [95].

If it is not proximity, then the question is what really motivates local actors (i.e., residents, their civil society organizations, and the representatives in public administration). An alternative proposed for proximity or the narrow interpretation of backyard-protective reactions is “*identity of place*” or “*place attachment*” [96–98]. These are both concepts in which the focus of explanation shifts away from the physical aspects of the development toward the socially constructed, symbolic attributes of places. It is essential how the identity of a place is interpreted by members of the community (not primarily restricted to residents) in order to “fit” with development proposals. Several empirical findings support the significance of the link between responses to

facilities and place attachment and community commitment. A well-known example is hydropower [99], but recent studies on wind power have also confirmed the importance of this relationship, among residents and local communities [57]. Empirical support for the significance of place identity is presented in a study in two coastal communities in Wales that are confronted with the same nearshore wind developments (North Hoyle and Gwynt y Mor). In one community, clear correlations between place attachment and landscape-identity evaluations about the wind farms were found, and in the other there was not such a relation (“Fence in the bay”: $r = 0.20$). Accordingly, in the community where people saw this relation between their place attachment and the landscape impact of the wind farm, there also was a correlation with oppositional behavior ($r = 0.20$) and none in the other community [98, p. 276].

In the concept of place attachment, there is a prominence of social and physical elements, whereby the latter are connected to landscape characteristics in particular [100]. This follows the general finding that landscape quality is by far the most important attribute of attitudes toward wind power projects. In environmental conflicts about siting wind power schemes, in the commonsense approach, the wind turbines are the object of the conflict; however, in fact, the physical space is the object of the conflict and not only the location of the conflict (section “[Landscape/Seascape](#)”).

The meaning of the concept of “community” can also be based on the feeling attached to the physical as well as the social attributes of a place. For example, place attachment can be high among tourists coming to the community, whereas in the community itself people care for this identity because the attraction is considered economically valuable. Another important group combining physical and social characteristics of the community are the members of organizations that work on the preservation and conservation of certain areas. The conflict over renewable energy facilities has often a character what has been labelled as “green on green” [42], with organizations that belong to the environmental movement at both sides of the conflicts. This may be related to the time scope of the environmental issues: some impacts become immediately manifest, in particular any visual change in the landscape, whereas others have a long-term risk character, among these most effects of climate change.

As many environmental organizations have primarily emerged because of landscape and nature protection values, their natural reaction to wind farms also concerns the potential threat to these values. Their identification will usually be strong because of values of landscape and ecology, but the identification will also be rooted in historic links to a place or in general ideas about the character of areas, e.g., in terms of “rural” [101, 102]. Studies also showed that this identification with places does not necessarily reflect negatively on the development of wind power (see section “[Landscape/Seascape](#)”) and the clearest example of this is the fact that farmers often have positive connotation with “harvesting” the wind on their land because they are used to associating their land with production. On the other hand, in some cases, proposed wind developments are the origin of new emerging landscape protection organizations [103].

Benefits

The most problematic aspect of community acceptance is caused by the fact that in many countries the framework of sociopolitical acceptance leads to projects that are initiated by actors who are outsiders to the community. First, this model neglects the potential of support for wind power initiatives that originates from within the community. Second, the community and the places valued by the community are used for interventions that do not create benefits for the community. Some studies suggest that local financial benefits “are critical to the acceptance of new turbines” [91, p. 507], or, put in a more prudent way, “renewable energy projects can become more locally divisive and controversial if benefits are not generally shared among local people” [33, p. 499]. Clearly, the benefits are shared when community members are shareholders in the wind farm, and the conflict of private vs. public interest is effectively mitigated.

Thus, in all cases of wind farms initiated by developers outside of the community, securing benefits for the community represent a significant problem. Nevertheless, community benefits can also be defined more broadly than direct financial gain, such as benefits-in-kind (provision of assets by the developer, e.g., parking facilities, children’s playground, educational material for schools, footpath, etc.) or investments

in environmental mitigation or enhancement [104]. As long as community members perceive the local benefits as valuable, they can be considered substantial, for example when they concern involvement of local small companies and employees in construction, maintenance, and management of the wind farms. Contribution of renewables is often suggested to positively affect local sustainability in a general sense. In the same way that usually the environmental dimension of sustainability is emphasized [105], the local dimensions of the social – inclusion, governance – and economical dimension – development, efficiency, employment – may also be promoted by a wind farm [106]. Several studies have shown indications that wind farm developments have indeed furthered local employment and the local or regional economy. However, these results remain ambiguous: the effects cannot be considered substantial [107] or the benefits are not experienced in the community but more dispersed, so general employment benefits are found at the national scale [108]. Special cases of community economies that may benefit from wind power projects can be found on islands, where wind power may help to reinforce the sustainability of the community [78, 109]. The general picture of local sustainability through wind energy is ambiguous, without any notable general tendency. This is reflected in the discourses around decision making on wind projects in the community. A clear objector's discourse has been recognized in a discourse study focussing on presumed negative local economic impact [68]. A general framework to analyze the contribution of renewable energy deployment to local sustainability has been proposed by Del Rio and Burguillo [106].

Some aspects should still be mentioned, as in certain cases they affect the acceptance of communities. These are farming and tourism, both economic sectors that may be considered land uses that compete with wind power deployment. Concerning the contribution of wind power to local sustainability, there is some evidence that rural areas may benefit because farmers are able to continue their business with a new resource to harvest from their land. As a matter of fact, this kind of positive acceptance is clear in most countries as many private investments in wind have been made by farmers on their own land. The picture for tourism is generally less positive. Community acceptance is adversely affected by the perceptions that clients, the tourists, would not like

turbines. Communities that heavily dependent on tourism, and in particular the stakeholders with interest in those activities, tend to be reluctant to allow renewable energy infrastructure – including wind turbines in the landscape – because it may easily affect the character of the area that is considered the primary attraction for tourists [110]. Several studies on the cost and benefits of wind power for tourism show ambiguous results, highly determined by the kind of landscape and the kind of tourism [111, 112]. An interesting topic is in particular the relation between wind power and tourism on island communities and also of nearshore wind power and tourism in coastal areas generally [113]. It is not always clear whether the tourists really do not like the turbines or that this is primarily untested fear of the actors in the tourist sector. In some cases, there is evidence that wind turbines may add to the tourist value of remote communities as they are seen as emphasizing the character of the location [93].

Fairness and Trust

Having local benefits that balance the burden that the community has to bear – at least in their own view – is a question of equity and fairness that is typical of environmental conflict and decision making on infrastructure [89, 90]. These are questions related to distributive justice (How are costs and benefits shared?), fairness of process (Is there a fair decision-making process that provides all relevant stakeholders with an opportunity to participate?), and trust (Are the information and the intentions of the investors and actors from outside the community trusted?) [52, 114].

These questions are strongly affected by the degree of collaborative decision making (Fig. 3: 3). In cases where the wind farm project clearly generates community benefits, the problems with community acceptance will remain limited. This often holds for different sorts of community ownership, but this will not be sufficient, in particular when the benefits are unequally distributed among the members of the community. Furthermore, the dimension of community involvement may also create feelings of (un)fairness if community members feel that the project is imposed on the community.

The “utility wind farm,” or any other project initiated outside the community, is clearly the model with

the greatest probability of meeting distrust [115]. In particular in policies that favor developments by large companies, there is a tendency to include some prescribed compensation to generate community benefits. However, this is a dangerous strategy, as the intrinsic motivation of the community members with a positive acceptance can easily be undermined [116]. Some practices of compensation in wind power cases show that the introduction of community benefits in case of a wind farm built by a community outsider may reinforce feelings of mistrust and it can easily be interpreted as an effort to bribe the opposition. Although community benefits, financial or otherwise, are presented as means of creating greater equity, experiences with this kind of compensation of local burden to the community show that “even the most well intentioned developers may have a hard time earning the trust of local communities” [117].

Market Acceptance

Social acceptance can also be interpreted as market acceptance or as the process of market adoption of an innovative STS. The literature on diffusion of innovation explains that the adoption of innovative products by consumers takes place through a communication and negotiation process between individual adopters and their environment [118]. A technology like wind power is bound to infrastructure, such as the power grid, which makes it inherently more complex to diffuse this innovation than other products. However, using the insights from this literature, it is helpful to study market acceptance at the consumer level.

Consumers can, and in successful examples of implementation do, play a role as investors. They have participated in cooperatives or other civil initiatives – “*Bürgerinitiative*” in German – to establish wind farms and thus became shareholders. There is a wide range of motives for such initiatives, but a strong driving force behind this development were various grass-roots initiatives (Fig. 3) based on environmental concern and the willingness to be involved in the development of alternatives to conventional and nuclear power generation [88, 119]. In the meantime, such civil initiatives have evolved into new energy companies with an inherent high degree of acceptance of renewables energy.

The alternative to investments by civilians in wind power is the choice in their role as consumers to subscribe to any kind of “green power” [120]. Such labels are in fact a type of marketing strategy used by energy companies, generating power through a mixture of sustainable energy generation [30, 121]. Most of these schemes include wind power, and consumers have the opportunity to switch to renewable energy supply without being actually involved in the physical generation. The “green power” schemes are probably the area where market adoption is rather isolated from the broader social acceptance picture. A study from Sweden shows that rejection or acceptance of green power does not correlate with proximity to renewable power facilities [69]. Nevertheless, if consumers would demand increased amounts of green power, wind farms would still need to be sited to supply this demand. The separation between (physical) supply and demand is inherent in the concept of green power marketing and trading, and it is a continuation of existing institutions in the power sector. The model is strongly favored by the incumbents in the energy sector, similar to the support they provide to RPS systems. This might aggravate the problem of social acceptance, for example if there is high green power demand, but the energy companies are not able to generate sufficient sociopolitical and community acceptance for building the corresponding wind power production infrastructure. Extensive international green power trade, that might also be favored by an internationally (e.g., EU wide) operating green certificates system, something that is strongly supported by Eurelectric – the European umbrella organization of the power supply sector – would reinforce such problems [79]. It seems questionable whether residents of a country with large wind resources, such as Scotland, would be willing to bear the burden of the export of green power generated by many large wind farms. It would become a market model that would provide a “free rider” escape for countries that scale-up the original “backyard” concept to their entire nation [122, p. 209] in which they do not opt to produce a sufficient amount of renewable energy within their own country needed for making the transition to a sustainable energy supply.

In a wider understanding of market acceptance, the focus is not just on consumers but also on investors and the structure of the economy. Examples of studies on

market acceptance are, for example, the “translation” of market stimulation and the effect of policy instruments as applied in Europe, in different markets such as the US. Bolinger [123] emphasizes the importance of the use of instruments that rapidly develop the domestic market for wind technology, in combination with incentives for using power technology that is manufactured locally. As the studies comparing the policies applied by different US states show, the socio-political acceptance of such measures is very different [49–51]. Equally important is the acceptance on new but rapidly developing markets like China. There the relation between the development of a domestic market and the development of a wind turbine industry has been reversed. The market is working under a strong centrally directed development regime, including the obligation to use wind turbines built in China, and moreover strong institutional impediments exist with regards the effective inclusion of regional and local identity in the wind developments. Together with the fact that the general market is determined by set prices in cases of noncompetitive contract wind farms, or in underbidding in cases of competitive concession projects, many wind farms are not embedded in the regional or local economy and they are often qualitatively low developments suffering from poor economic performance [124].

With all kinds of market systems, there is also the issue of intra-firm acceptance of renewable energy innovation. Numerous examples show that large firms are subject to path dependency when it comes to their investment behavior. For the general picture of social acceptance, early market formation is an essential contribution to building the legitimacy of renewable energy innovation as a whole [125, 126]. The issue of wind power implementation is also connected with the development of a wind power technology industry, as the existence and progressive development of a domestic market for application is the key to the development of this new branch of industry [127].

The success story countries developed a strong domestic market without large investments from existing energy companies. How social acceptance is constructed within existing energy companies would be an issue well worth studying. This could very well be linked to some of the research on cognitive barriers within firms with regard to taking up environmental

and sustainability issues [128]. Many of the energy companies still own and manage significant parts of the grid; even though in some countries, the grid management is subdivided in regional monopolies and is separated from power generation. The role, position, and objectives of existing energy companies are a strong manifestation of market acceptance. In the Netherlands, the energy companies held a gatekeeper role and decided about the remuneration and grid access for more than a decade, whereas the German government forced power companies to accept competition from third parties [88]. The latter factor is an essential feature of the success of the German “feed-in” legislation and the Renewable Energies Act which effectively stimulated third parties to invest in wind turbines [119]. It is clear, particularly from the position of the energy companies toward the REFIT and RPS support schemes, that they prefer systems that enhance their position in the market over the introduction of systems that stimulate third-party investments in small-scale renewable power generation [73, 78]. Moreover, there is a link with sociopolitical acceptance, because the incumbent energy companies are influential stakeholders in the development of energy policies and they try to use their influence in the crucial political decisions about the design of financial procurement systems, the access to the grid for other investors in renewable energy systems, and the introduction of green certificates trading systems [79].

Significant Attributes Connected to Place Identity

Numerous potential attributes of wind power application are shown in Table 2. Most of these with high relevance for sociopolitical acceptance are connected to issues concerning governance and institutional change, or they are general attributes of the wind power technology, such as its variable supply character. Some attributes come to the fore on the level of general application of wind power as well as in any specific case of building a wind farm. The way they show up in these cases, however, is very different. These attributes concern the impact that building wind turbines has on the environment, and these impacts are not primarily determined by the wind turbines itself or by the wind farm design. The impact is primarily determined by the existing local environment before establishing the

development. Three categories of impact can be distinguished that are discussed in almost any case: landscape, wildlife, and annoyance.

Landscape/Seascape

The attribute “landscape” is by far the most significant in social acceptance. Attitude studies in the 1980s and 1990s already indicated that the discussions on the desirability of large-scale application of wind power focus upon the degree to which wind turbines can be integrated in the landscape [3, 4, 41]. In most policy documents and in approaches of developers, this weak spot is recognized, but the issue is poorly understood. In the realms of policy and developers, it is usually narrowed down to “visual impact” [39, 129]. Two approaches can be recognized. In the first approach, studies try to develop instruments for visual impact assessment, and these tend to focus only on the visibility as the determining factor [130]. The second approach tries to include assessments of landscape by the public, but this approach is also meant to find generally applicable parameters of visual impact in order to quantify the corresponding impact [131].

With regards the issue of social acceptance, both approaches fail to address the essence of the landscape issue. The inevitability of the highly visual character of wind power is reason of the importance of the landscape issue, but the visibility is not the problematic impact itself. The visibility is making the landscape the most important – salient – attribute of wind power acceptance because it unifies three kinds of strongly varying elements. Landscape assessment of wind power varies, because:

1. The landscape impact varies widely among cases, as the character of the landscape is very different at the locations.
2. Within each landscape, there are many distinguishable elements that may differently be affected.
3. The valuation of the impact shows a wide variety among individuals: some can positively value landscape elements, whereas other individuals value the same element negatively.

Furthermore, the third factor of subjective valuation is also connected to the geography of the case, as local communities may differently assess similar

landscapes because the identity of their community, including the landscape surrounding it, is historically and culturally determined. This is what is making the landscape issue highly subjective, and impossible to quantify objectively.

The notion of “visual impact” almost automatically classifies the impact on the landscape as negative. This strong simplification can also be found among the minorities that oppose to wind power in general. In individual attitudes, the beliefs and valuation concerning landscape are the strongest determinants of attitudes toward the energy source wind power [41, 52]. A clear example of this is the sociocultural approach by Brittan [132] who uses a terminology that classifies wind turbines as “alien” to the landscape, determining their unacceptability. He compares wind turbines to weeds, “invaders” offending aesthetic values in a field of native flora: “To the extent that standardized machines are plunked down in a standardized way, then no matter who owns them, the local character of the community is thereby weakened if not also destroyed” [132, p. 178]. This is clearly the philosophy of a strong general opponent of wind power, but the quote also underlines the significance of landscape and place identity for this attitude (see section “Public Opinion”).

In the late 1980s, the first studies that tried to establish the significance of arguments that inform attitudes to wind energy already showed that these attitudes primarily root in values concerning landscape [4, 41]. Ever since, public perception studies as well as acceptability case studies show that attitudes mainly reflect the assessment of the acceptability of turbines in terms of congruency between turbines and the type of landscape in which they are sited [133–135]. Despite the enormous variation that already exists between “typical” national or regional landscapes, studies in different countries have shown that it is the type of landscape in which the turbines are sited that is the most determining factor in acceptance of wind power. This holds for onshore as well as offshore [136, 137]. The idea that acceptability offshore will be higher and easier to achieve assumes that offshore wind power schemes would avoid the issue of destroying the landscape, which is again based on the simplistic idea that visibility determines landscape impact. Several recent studies show that the idea that building wind farms “over the sea and far away” would solve the social

acceptance issue is naïve [138]. Landscapes may also be “seascapes” and these are also determining public acceptability, in particular since most offshore schemes are still nearshore.

Several studies of nearshore wind power have shown evidence that social acceptance is inevitably a topic, and the crucial factor of the impact on the seascape has indeed come to the fore [139, 140]. Table 3 shows the impact of visual factors of a large proposed nearshore wind farm on the attitudes toward wind power development in the region. The variance in the acceptability could largely be explained by these visual factors, but in fact there were only two variables that described the character of the landscape that were important. Landscapes with ecological value are clearly negatively associated with siting wind turbines, but it is important to see that they are very positively associated with areas already in use for economic activities. All visual factors concerning the turbines and the way the wind farm would be designed did not significantly add explained variance. Hence, it is the qualification of the landscape that is the key to visual impact. Some cases show that the identification to seascapes with high iconic value – e.g., Nantucket Sound MA in the US and the WaddenSea in

the Netherlands – has produced strong oppositional movements and eventually failing large wind farm proposals [44, 141, 142]. These studies also show the significance of the concept of “place identity,” as the values attached to those seascapes are strongly associated with cultural heritage and ecological values, giving these conflicts a typical “green on green” character [42].

The question that remains is how to handle the landscape/seascape issue. Any assessment of the congruence between local landscapes and a wind power scheme may be very subjective indeed, but nevertheless it is the number one issue that should be addressed in decision making. There are two strategies to handle this issue that usually enhance the conflict, but that are frequently applied by developers and authorities. The first strategy is trying to “objectify” the assessments by efforts to develop instruments for the objective assessment of visual impact, which was, for example, tried with an instrument of “viewshed simulations” applied in the Nantucket Sound case (“Cape Wind”) [141]. A second strategy is to hire one of more “experts” to finally decide what the visual impact is and this strategy was applied in the WaddenSea case by calling in the official state’s architect [142]. In both cases, the strategies failed because these methods try to abstract the essence of the identity of the landscape, which is that it is primarily the values that determine this place identity.

For any decision that addresses the identity of the landscape, it is crucial to apply a method that assesses landscape quality interaction with the new elements – the wind turbine – as it is actually perceived by the relevant viewers. By definition, landscape is the part of the environment that is the human habitat as it is perceived and understood through the medium of our perceptions [143]. Apparently, most developers and authorities alike do not like it, but these perceptions of landscapes are rooted in community values and history, and the only way to deal with them is to deal with the people holding these values. It is crucial to apply methods that assess landscape quality interaction as it is actually perceived by the viewers, because the value of the landscape is “in the eye of the beholder” [144, 145]. Together with the strong influence of the landscape issue on people’s acceptability judgments, these observations all underline why collaborative planning strategies – see Fig. 3 – including cooperative and participatory decision making are crucial.

Wind Power: Basic Challenge Concerning Social Acceptance. Table 3 Support for wind power development in the entire region, explained by landscape type factors and design factors [52]; standardized regression coefficients (β)

Landscape/design factor	β	t	$p < 0.01$
Landscape I character: economically applied area	0.45	10.5	+
Landscape II character: nature, ecological value	−0.28	−7.9	+
Landscape III character: residential use	0.06	1.53	$p > 0.1$
Landscape IV character: offshore (North Sea)	−0.03	−0.14	$p > 0.1$
Wind farm design I large/small wind farms	0.01	0.33	$p > 0.1$
Wind farm design II small/tall turbines	0.01	0.23	$p > 0.1$
Wind farm design III small/large numbers	0.04	1.10	$p > 0.1$

$N = 535$; $R = 0.66$; $R^2 = 0.44$

This applies to onshore wind, but also to offshore [140] although the communities that are involved will often be different. For offshore wind farms a new practice of spatial planning at sea is emerging, and similar to planning on land, the involvement of crucial stakeholders is essential [146, 147]. Most of these are representing competitive uses of space at sea and coastal land uses that can be affected by near-shore wind development. To name a few: tourism, nature wildlife, shipping tracks, fisheries, oil and gas exploitation, national defence, etc. Some experimental involvements of vested interests at sea are currently investigated, in particular the co-development of wind power with new ways of marine culture and commercial fisheries [148].

Wildlife

Nature and landscape are strongly connected in perceptions. The strong commitment and tacit knowledge about the ecological qualities of the landscape are also shaping landscape perceptions within communities. This results in the “green-on-green” character of many conflicts about wind power schemes as organizations founded for the protection of natural values and wildlife are often involved in questioning wind power developments [42].

Similar to the attribute “landscape,” the fundamental factors that determine the actor’s assessment of the attribute “nature and wildlife” are the visibility and the identity of the location in the eyes of the community members. The fact that the casualties of birds are so visible is making the phenomenon inevitable for policy makers, developers, and local decision makers. In terms of attitudes, with regards bird casualties, we are dealing with beliefs about a “salient” attribute ([28]; see section “Attitude Objects and Attributes”). Efforts to assess the factual impact on avian wildlife are still not definitive. The impact is heavily dependent on what kind of species of birds (and similarly also bats) could be affected. Furthermore, the acceptability is as always not only depending on the factual impact of wind turbines on wildlife but it is also depending on similar impact of the alternatives to wind power [149]. Hence, there is no doubt that there is some impact on avian wildlife, and even serious adverse consequence on some species, but in fact the

impact of alternatives, in particular coal and nuclear power on wildlife (including the impact of power lines) is probably much stronger. The impact of hydropower is not primarily on avian wildlife, but its impact on fish and river ecosystems, in general, seems very high, but how should this be compared to the impact of wind turbines on some species? Because of the high visibility, wind power is vulnerable to a more negative profile than hydropower has in regard to its impact on fish. The high visibility of wind turbines and of eventual casualties is making this impact a serious problem with regard to the acceptance of wind turbines at certain locations.

The impact on birds is mainly connected to the location. The presence of birds is obviously geographically determined – due to breeding grounds, migration routes, etc. – and these are fairly objective factors, but the social valuation of the impact depends heavily on how the affected species are valued. These are species that are already rare and endangered, but also species that enjoy a valued status as icons of certain places, for example the sea eagle, a bird with high iconic value that has spurred discussion about the closing of an existing wind farm in Norway [150]. The strongest valuation is again connected to the characteristic presence of certain species or of entire ecosystems in which these species are essential elements. The presence of such species can add much to the identity of certain locations and then it will strongly affect the acceptability in the eyes of the communities that strongly identify with the place, such as nature-protection organizations.

In the collaborative decision making, the discussion may be focused upon accepting the building of the wind turbines at the site or to select another site. It may also concern the way of siting the wind farm, whereby specific characteristics of the living patterns of the birds can be used to redesign the wind farm [151]. The impact on wildlife is not the strongest factor in determining attitudes to wind farm developments, but the issue is closely linked to question of place identity and in particular landscape identity (section “Identity of Place”). Hence, for social acceptance it is very important that the issue is addressed in the collaborative decision-making process by means of participation of the organizations representing the interest of nature and wildlife.

Annoyance: Noise

The local impact of wind turbines is often described in terms of annoyance. For example, the visibility of the turbines is sometimes referred to as visual annoyance or intrusion. As described in section “[Landscape/Seascape](#),” this description is mainly determined by valuation. The most serious sources of annoyance are shadow flicker and the sound produced by wind turbines. The valuation of that sound is usually referred to as “noise,” a term with a clearly negative connotation. Shadow flicker is very severe indeed, but it can be avoided because the places where the moving shadows appear can easily be calculated [152]. The perception and experience of noise, however, is a very complicated matter.

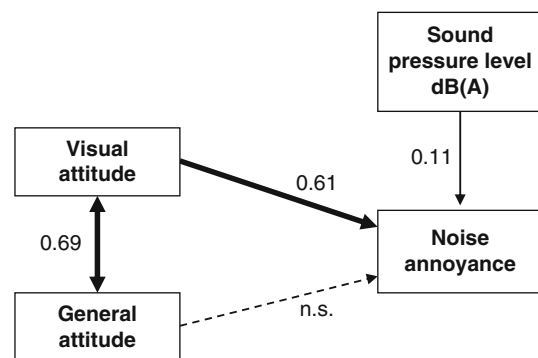
Noise has been addressed from the early 1980s onward as a serious issue concerning wind turbines. Technology development has reduced the production of sound by wind turbines, but simultaneously the size of the turbines has increased. In decision making about the siting of turbines, noise can play a significant role. In the collaborative process, the issue of sound produced by the turbines should be taken seriously. However, this does not mean that the expectations about noise annoyance are primarily determined by the sound of the turbines. The first study on the relation between the actual sound pressure produced by wind turbines and annoyance was carried out in the early 1990s when the number of people living close enough to wind turbines to be seriously affected was rather limited. This study in three countries found that the level of annoyance among surrounding residents was stronger affected by attitudes to the wind turbines than by the factual sound pressure [153].

This phenomenon has been replicated in a more sophisticated noise annoyance study. It combined on-site physical measurements of sound pressure with adequate measurements of all relevant variables to annoyance among residents on several locations in Sweden. In that study, the strongest influence on annoyance was from two attitude variables concerning wind turbines in general and to the visual impact of the wind turbines [154, p. 484]. In a second study, these relations were specified in a model and tested, and the result is shown in Fig. 4 [155]. The figure shows that there is a positive correlation between dB(A) sound pressure levels and annoyance, but noise annoyance is

much stronger related to the visual attitude. This visual attitude has been measured with indicators as the impact on the landscape, a judgment in terms of beautiful or ugly, and a judgment about natural versus unnatural. As usual, this attitude is in turn strongly related to the general attitude to wind turbines, measured in terms of general opinion, efficient power generation, and necessity. Figure 4 presents the results for those residents that could see at least one wind turbine. There were no differences between rural and built-up areas, but the relations became weaker on locations with hilly or rocky terrain and among people that could not see any of the wind turbines. The study shows that even for noise annoyance, which is about the sound that is produced by the turbines, the landscape issue is still a more important attribute than the sound itself. This finding reinforces the overall dominant significance of landscape for acceptance of wind turbines.

Future Directions

The issue of social acceptance of wind power will come to the fore even more prominent in the coming decades, but the character of acceptance issues in the three dimensions will probably change. In the coming decades, the problems concerning the institutional constraints for the emergence of a new STS in its early adoption phase will gradually shift toward the social



Wind Power: Basic Challenge Concerning Social Acceptance. Figure 4

Structural equation model [155, p. 384] presenting strength of paths explaining noise annoyance for residents who could see at least one turbine ($N = 843$)

acceptance of interconnected sustainable energy systems in which wind power is only one of the substantial components. It will no longer be an issue of introducing a new kind of energy supply, but of implementation of sufficient capacity of renewable sources required to establish a full transition toward sustainable energy supply. In all three dimensions of social acceptance, new problematic issues will emerge because of the enormous requirements in the amount of space needed to establish that capacity. This is considerably underestimated in policy – as well as the rest of the society – in particular with regards the power–space ratio in relation to the energy consumption in developed countries. This ratio shows the impressive amounts of land-use change needed [122].

The current awareness of the required space for sustainable energy supply is still limited and, hence, the consequences in terms of landscape occupation and potential environmental conflicts are neither fully recognized. One of the major changes that is recognized though, is the potential rapid development offshore. This constitutes a major new challenge, but as indicated – section “[Landscape/Seascape](#)” – the issue of social acceptance may be different there because of other competing spatial claims and key stakeholders, but this definitely does not result in less complexity in decision making. Again, a collaborative approach to decision making is needed to handle these complexities in a way that furthers acceptance among the crucial actors.

The essential role for collaborative planning approaches to wind power emphasizes the role of communities. This line of practice will only become more important because of the number of wind power projects and therefore the number of communities that need to be involved. There should be no misunderstanding though, that the complexities of the realities of community can easily be underestimated. Although most acceptance studies reveal that the role of communities is essential, there should be no presumptions about communities that they are always benign toward sustainability [156] and their willingness to further the development of renewable and sustainable energy cannot be taken for granted. A comparison of different cases of community energy has shown that community support for wind power projects, even if they are rooted in parts of the community, is definitely not self-evident.

This analysis concludes that “trust between local people and groups that take projects forward is part of the package of conditions which can help projects work and for local people to feel positive about getting involved and about process of project development” [157, p. 2662]. However, there are no simplistic prescriptions of “what works” and community wind power projects cannot simply be replicated from place to place. Again, the essence of the implementation of a new STS in a community is that understanding the social context of innovation and technology diffusion is just as important as its technical dimensions. Studies that enhance our understanding of how to embed renewable energy supply in communities still have a short history, and the understanding is still limited.

The most significant aspect of collaborative decision making about and community acceptance of wind power projects is the fit to the identity of the landscape in the eyes of the community members. Within the sociopolitical acceptance dimension, there is a persistent preference for efforts to try to avoid the far-reaching subjectivity of “the eyes of the beholder” [144]. The efforts to develop technological instruments to objectively assess the “visual impact” will not enhance trust, as they are not helpful to address the valuation of landscape identity. A more fruitful direction for the analysis of the landscape issue with newly developed technological instruments is where such techniques are applied to support the collaborative planning process of wind power schemes. Recently, some studies have shown considerable progress in that direction [158, 159].

The main question of social acceptance will remain how to build sociopolitical and market acceptance for the collaborative way of planning and decision making that is needed. This key to the large number of positive investment and space-making decisions will even become more important because wind power is becoming increasingly part of an all-embracing STS of integrated sustainable energy supply. The new power supply system will have to integrate growing numbers of distributed generation units [26, 160]. This STS may include several components, such as solar PV, small-scale CHP, tidal, geothermal, and hydropower, and all of these are subject to questions with regard social acceptance [161]. Furthermore, the increasing amount of capacity with a variable supply character is adding

new challenges to the power grid, which in the current situation is considered storage capacity for sources like wind power [162]. For sociopolitical acceptance, this development asks for willingness to invest in the development of a new STS based on a power grid that can handle the increasing tendency of distributed generation and decentralized optimization [163].

The integration of all the components of sustainable power generation, including the mutual fine-tuning and optimization of local supply and demand and possibly with the introduction of local storage capacity – e.g., electric vehicles [164] – must be embedded in a “smart grid” [165]. All the new technologies needed for this integration will become subject to acceptance decisions, as incumbents in the existing power sector should institutionally support them in policy, and they need willingness to investments by many market actors, including the acceptance by consumers. The issue of community acceptance will come to the fore more prominent as well, as we are dealing with more distributed generation that by nature depends largely on decisions on that level. The future “smart grid” will probably be something like “a network of integrated micro-grids that can monitor and heal itself” [165, p. 570] in which the micro-grids are in fact connecting communities of consumers that partly have become producers themselves – or coproducers – and mutual suppliers. All questions of (co-)ownership and involvement in decision making that have come to the fore as crucial in wind power development will become more complex. They become embedded in questions about ownership and about who will be in control of all the facilities that are part of the new supply system, which in particular includes the smart meters and smart regulating devices that will be the hubs of the smart micro-grids. The social acceptance of wind power will become embedded in the acceptance of all kinds of decisions about this future STS of sustainable energy supply and demand.

Bibliography

Primary Literature

1. Carlman I (1982) Wind energy potential in Sweden: the importance of non-technical factors. In: Fourth international symposium wind energy systems, Stockholm, pp 335–348, 21–24 Sept 1982
2. Bosley P, Bosley K (1988) Public acceptability of California's wind energy development: three studies. *Wind Eng* 12:311–318
3. Wolsink M (1986) Public acceptance of large WECS in the Netherlands. In: European wind energy conference, 1986, vol II. Raguzzi, Rome, pp 587–592
4. Thayer RL, Freeman C (1987) Public perceptions of a wind energy landscape. *Landscape Urban Plan* 14:379–398
5. Carlman I (1984) The views of politicians and decision-makers on planning for the use of wind power in Sweden. In: European wind energy conference, Hamburg, pp 339–343, 22–26 Oct 1984
6. Wolsink M (1987) Wind power for the electricity supply of houses. *Neth J House Environ Res* 2:95–214
7. Aitken M (2010) Why we still don't understand the social aspects of wind power: a critique of key assumptions within the literature. *Energy Policy* 38:1134–1841
8. Wüstenhagen R, Wolsink M, Bürer MJ (2007) Social acceptance of renewable energy innovation: an introduction to the concept. *Energy Policy* 35:2683–2691
9. Bijker WE, Law J (eds) (1992) *Shaping technology/building society: studies in sociotechnical change*. MIT Press, Cambridge, MA
10. Geels FW (2004) From sectoral systems of innovation to socio-technical systems – insights about dynamics and change from sociology and institutional theory. *Res Policy* 33:897–920
11. Hekkert M, Suurs RAA, Negro SO, Kuhlmann S, Smits REHM (2007) Functions of innovation systems: a new approach for analysing technological change. *Technol Forecast Soc* 74:413–432
12. Jacobsson S, Johnson A (2000) The diffusion of renewable energy technology: an analytical framework and key issues for research. *Energy Policy* 28:625–640
13. Kamp LM, Smit REHM, Andriess CD (2004) Notions on learning applied to wind turbine development in the Netherlands and Denmark. *Energy Policy* 32:1625–1637
14. North D (1990) *Institutions, institutional change and economic performance*. Cambridge University Press, Cambridge
15. Thelen K (1999) Historical institutionalism in comparative politics. *Annu Rev Polit Sci* 2:369–404
16. Scott WR (1987) The adolescence of institutional theory. *Admin Sci Quart* 32: 117, 493–511
17. Powell WW, DiMaggio PJ (eds) (1991) *The new institutionalism in organizational analysis*. Chicago University Press, Chicago
18. Hoogwijk M, de Vries B, Turkenburg W (2004) Assessment of the global and regional geographical technical and economic potential of on-shore of wind energy. *Energy Econ* 26:889–919
19. Yue C-D, Min-How Y (2009) Exploring the potential of wind energy for a coastal state. *Energy Policy* 37:3925–3940
20. Sovacool BK (2009) Rejecting renewables: the socio-technical impediments to renewable electricity in the United States. *Energy Policy* 37:4500–4513
21. Wolsink M, Breukers S (2010) Contrasting the core beliefs regarding the effective implementation of wind power

- international study stakeholder perspectives. *J Environ Plann Man* 53:535–558
22. Jessup B (2010) Plural and hybrid environmental values: a discourse analysis of the wind energy conflict in Australia and the United Kingdom. *Environ Polit* 19:21–44
 23. Wolsink M (2000) Wind power and the NIMBY-myth: institutional capacity and the limited significance of public support. *Renew Energ* 21:49–64
 24. Baumeister RF, Leary MR (1995) The need to belong – desire for interpersonal attachments as a fundamental human motivation. *Psychol Bull* 117:497–529
 25. Rhodes RAW (1996) The new governance: governing without government. *Polit Stud* 44:652–667
 26. Ackermann T, Andersson G, Söder L (2001) Distributed generation: a definition. *Electr Pow Syst Res* 57:105–204
 27. Devine-Wright P, Devine-Wright H (2006) Social representations of intermittency and the shaping of public support for wind energy in the UK. *Int J Glob Energ* 25:243–256
 28. Ajzen I (1991) The theory of planned behaviour. *Organ Behav Hum Dec* 50:179–211
 29. Ajzen I (2001) Nature and operation of attitudes. *Annu Rev Psychol* 52:27–58
 30. Bang HK, Ellinger AE, Hadjimarcou J, Traichal PA (2000) Consumer concern, knowledge, belief, and attitudes toward renewable energy: an application of the reasoned action theory. *Psychol Market* 17:449–468
 31. Dinica V (2008) Initiating a sustained diffusion of wind power: the role of public-private partnerships in Spain. *Energy Policy* 36:3562–3571
 32. Möller B (2006) Changing wind-power landscapes: regional assessment of visual impact on land use and population in Northern Jutland, Denmark. *Appl Energ* 83:477–494
 33. Walker G, Devine-Wright P (2008) Community renewable energy: what should it mean? *Energy Policy* 36:497–450
 34. Bell D, Gray T, Haggett C (2005) The ‘social gap’ in wind farm siting decisions: explanations and policy responses. *Environ Polit* 14:460–477
 35. Toke D (2005) Explaining wind power planning outcomes: some findings from a study in England and Wales. *Energy Policy* 33:1527–1539
 36. European Wind Energy Association (EWEA) (2009) Wind energy: the facts. A guide to the technology, economics and future of wind power. Earthscan, London/Sterling
 37. Freudenburg WR, Pastor SK (1992) Nimby and lulus: stalking the syndromes. *J Soc Issues* 48:39–61
 38. Devine-Wright P (2005) Beyond NIMBYism: towards an integrated framework for understanding public perceptions of wind energy. *Wind Energ* 8:125–139
 39. Kahn RD (2000) Siting struggles: the unique challenge of permitting renewable energy power plants. *Electr J* 13:21–33
 40. Wolsink M (2006) Invalid theory impedes our understanding: a critique on the persistence of the language of NIMBY. *T I Brit Geogr* 31:85–91
 41. Wolsink M (1989) Attitudes and expectancies about wind turbines and wind farms. *Wind Eng* 13:196–206
 42. Warren CR, Lumsden C, O’Dowd S, Birnie RV (2005) ‘Green on green’: public perceptions of wind power in Scotland and Ireland. *J Environ Plann Man* 48:853–875
 43. Jones CR, Eiser JR (2009) Identifying predictors of attitudes towards local onshore wind development with reference to an English case study. *Energy Policy* 37:4604–4614
 44. Kempton W, Firestone J, Lilley J, Rouleau T, Whitaker P (2005) The off-shore wind power debate. *Views Cape Cod. Coast Manage* 33:119–149
 45. Devine-Wright P (2011) From backyards to places: public engagement and the emplacement of renewable energy technologies. In: Devine-Wright P (ed) *Renewable energy and the public: from NIMBY to participation*. Earthscan, London
 46. Sjöberg L, Drottz-Sjöberg B (2001) Fairness, risk and risk tolerance in the siting of a nuclear waste repository. *J Risk Res* 4:75–101
 47. Wolsink M, Devilee J (2009) The motives for accepting or rejecting waste infrastructure facilities. Shifting the focus from the planners’ perspective to fairness and community commitment. *J Environ Plann Man* 52:217–236
 48. Oxford English Dictionary (2010) <http://dictionary.oed.com/entrance.dtl>. Accessed 15 Apr 2010
 49. Dear M (1992) Understanding and overcoming the NIMBY syndrome. *J Am Plann Assoc* 58:288–300
 50. Swofford J, Slattery M (2010) Public attitudes of wind energy in Texas: local communities in close proximity to wind farms and their effect on decision-making. *Energy Policy* 38:2508–2519
 51. Toke D, Breukers S, Wolsink M (2008) Wind power deployment outcomes: how can we account for the differences? *Ren Sust Energ Rev* 12:1129–1147
 52. Wolsink M (2007) Planning of renewables schemes. Deliberative and fair decision-making on landscape issues instead of reproachful accusations of non-cooperation. *Energy Policy* 35:2692–2704
 53. Healey P (1997) *Collaborative planning: shaping places in fragmented societies*. Macmillan, London
 54. Marques AC, Fuinhas JA (2011) Drivers promoting renewable energy: a dynamic panel approach. *Ren Sust Energ Rev* 15:1601–1608
 55. Bohn C, Lant C (2009) Welcoming the wind? Determinants of wind power development among U.S. states. *Prof Geogr* 61:87–100
 56. Menz FC, Vachon S (2006) The effectiveness of different policy regimes for promoting wind power: experiences from the states. *Energy Policy* 34:1786–1796
 57. Ohl C, Eichhorn M (2010) The mismatch between regional spatial planning for wind power development in Germany and national eligibility criteria for feed-in tariffs – a case study in West Saxony. *Land Use Policy* 27:243–254
 58. Breukers S, Wolsink M (2007) Wind energy policies in the Netherlands: institutional capacity-building for ecological modernisation. *Environ Polit* 16:92–112
 59. Heiman MK, Solomon BD (2004) Power to the people: electric utility restructuring and the commitment to renewable energy. *Ann Assoc Am Geogr* 94:94–116

60. Wilson EJ, Stephens J (2009) Wind deployment in the United States: states, resources, policy, and discourse. *Environ Sci Technol* 43:9063–9070
61. Sabatier PA (1998) The advocacy coalition framework: revisions and relevance for Europe. *J Eur Publ Pol* 5:98–130
62. Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. *Science* 211:453–458
63. Walker G (1995) Renewable energy and the public. *Land Use Policy* 12:49–59
64. Eurobarometer (2007) Energy technologies: knowledge, perception measures. European Commission, Brussels
65. Ansolabehere S, Konisky DM (2009) Public attitudes toward construction of new power plants. *Public Opin Quart* 73:566–577
66. Klick H, Smith ERAN (2010) Public understanding of and support for wind power in the United States. *Renew Energy* 35:1585–1591
67. Wolsink M (2007) Wind power implementation: the nature of public attitudes: equity and fairness instead of “backyard motives”. *Ren Sust Energy Rev* 11:1188–1207
68. Ellis G, Barry J, Robinson C (2007) Many ways to say ‘no’, different ways to say ‘yes’: applying q-methodology to understand public acceptance of wind farm proposals. *J Environ Plann Man* 50:517–551
69. Ek K (2005) Public and private attitudes towards “green” electricity: the case of Swedish wind power. *Energy Policy* 33:1677–1689
70. Clark CF, Kotchen MJ, Moore MR (2003) Internal and external influences on pro-environmental behavior: participation in a green electricity program. *J Environ Psychol* 23:237–246
71. Ellis G, Cowell R, Warren C, Strachan P, Szarka J (2009) Expanding wind power: a problem of planning or of perception? *Plann Theory Pract* 10:523–532
72. Muñoz M, Oschmann V, Tabara JD (2007) Harmonization of renewable electricity feed-in laws in the European Union. *Energy Policy* 35:3104–3114
73. Lauber V (2004) REFIT and RPS: options for a harmonised community framework. *Energy Policy* 32:1405–1414
74. Toke D (2007) Renewable financial support systems and cost-effectiveness. *J Cleaner Prod* 15:280–287
75. Jacobsson S, Lauber V (2006) The politics and policy of energy system transformation – explaining the German diffusion of renewable energy technology. *Energy Policy* 34:256–276
76. Mitchell C, Bauknecht D, Connor PM (2006) Effectiveness through risk reduction: a comparison of the renewable obligation in England and Wales and the feed-in system in Germany. *Energy Policy* 34:297–305
77. Krewitt W, Nitsch J (2003) The German Renewable Energy Sources Act. An investment into the future pays off already today. *Renew Energy* 28:533–542
78. Toke D (2005) Are green certificates the way forward for renewable energy? An evaluation of the UK’s Renewable obligation in the context of international comparisons. *Environ Plann C* 23:361–375
79. Jacobsson S, Bergek A, Finon D, Lauber V, Mitchell C, Toke D, Verbruggen A (2009) EU renewable energy support policy: faith or facts? *Energy Policy* 37:2143–2146
80. Toke D (2008) The EU renewables directive – what is the fuss about trading? *Energy Policy* 36:3001–3008
81. Cowell R (2007) Wind power and “the planning problem”: the experience of Wales. *Eur Environ* 17:291–306
82. Bergek A (2010) Levelling the playing field? The influence of national wind power planning instruments on conflicts of interests in a Swedish county. *Energy Policy* 38:2357–2369
83. Toke D (2005) Explaining wind power planning outcomes: some findings from a study in England and Wales. *Energy Policy* 33:1527–1539
84. Dimitropoulos A, Kontoleon A (2009) Assessing the determinants of local acceptability of wind-farm investment: a choice experiment in the Greek Aegean Islands. *Energy Policy* 37:1842–1854
85. Loring J, MacLaren (2007) Wind energy planning in England, Wales and Denmark: factors influencing project success. *Energy Policy* 35:2648–2660
86. Aitken M, McDonald S, Strachan P (2008) Locating ‘power’ in wind power planning processes: the (not so) influential role of local objectors. *J Environ Plann Man* 51:777–799
87. Söderholm P, Ek K, Petterson M (2007) Wind power development in Sweden: global policies and local obstacles. *Ren Sust Energy Rev* 11:365–400
88. Agterbosch S, Breukers S (2008) Socio-political embedding of onshore wind power in the Netherlands and North Rhine-Westphalia. *Technol Anal Strateg Man* 20:633–648
89. Mannarini T, Roccato M, Fedi A, Rovere A (2009) Six factors fostering protest: predicting participation in locally unwanted land uses movements. *Polit Psychol* 30:895–920
90. Wolsink M (2010) Contested environmental policy infrastructure: socio-political acceptance of renewable energy, water, and waste facilities. *Environ Impact Asses* 30:302–311
91. Devlin E (2005) Factors affecting public acceptance of wind turbines in Sweden. *Wind Eng* 29:503–511
92. Musall FD, Kuik O (2011) Local acceptance of renewable energy – A case study from southeast Germany. *Energy Policy* 39:3252–3260
93. Warren CR, McFadyen M (2010) Does community ownership affect public attitudes to wind energy? A case study from south-west Scotland. *Land Use Policy* 27:204–313
94. Hindmarsh R, Matthews C (2008) Deliberative speak at the turbine face: community engagement, wind farms, and renewable energy transitions. *J Environ Pol Plann* 10:217–232
95. Graham JB, Stephenson JR, Smith IJ (2009) Public perceptions of wind energy developments: case studies from New Zealand. *Energy Policy* 37:3348–3357
96. Wester-Herber M (2004) Underlying concerns in land-use conflicts – the role of place-identity in risk perception. *Environ Sci Pol* 7:109–116
97. Jackson JB (1994) A sense of place, a sense of time. Yale University Press, New Haven

98. Devine-Wright P, Howes Y (2010) Disruption to place attachment and the protection of restorative environments: a wind energy case study. *J Environ Psychol* 30:271–280
99. Vorkinn M, Riese H (2001) Environmental concern in a local context. The significance of place attachment. *Environ Behav* 33:249–263
100. Stedman RC (2002) Is it really just a social construction? The contribution of the physical environment to sense of place. *Soc Nat Res* 16:671–685
101. Woods M (2003) Conflicting environmental visions of the rural: windfarm development in Mid Wales. *Sociol Ruralis* 43:171–288
102. Bergmann A, Colombo S, Hanley N (2008) Rural versus urban preferences for renewable energy developments. *Ecol Econ* 65:616–625
103. Moragues-Faus AM, Ortiz-Miranda D (2010) Local mobilization against wind farm developments in Spanish rural areas: new actors in the regulation arena. *Energy Policy* 38:4232–4240
104. Cowell R, Bristow G, Munday M (2011) Acceptance, acceptability and environmental justice: the role of community benefits in wind energy development. *J Environ Plann Man* 54:539–557
105. Dincer I (2000) Renewable energy and sustainable development: a crucial review. *Ren Sust Energy Rev* 4:157–175
106. delRio P, Burguillo M (2008) Assessing the impact of renewable energy deployment on local sustainability: towards a theoretical framework. *Ren Sust Energy Rev* 12:1325–1344
107. Faulin J, Lera F, Pintor JM, García J (2006) The outlook for renewable energy in Navarre: an economic profile. *Energy Policy* 34:2201–2216
108. Hillebrand B, Buttermann H, Behringer J, Bleuerl M (2006) The expansion of renewable energies and employment effects in Germany. *Energy Policy* 34:3484–3494
109. Kaldellis JK, Zafirakis D (2007) Present situation and future prospects of electricity generation in Aegean Archipelago islands. *Energy Policy* 35:4623–4639
110. Dalton GJ, Lockington DA, Baldock TE (2008) A survey of tourist attitudes to renewable energy supply in Australian hotel accommodation. *Renew Energy* 33:2174–2185
111. Frantál B, Kunc J (2011) Wind turbines in tourism landscapes: Czech experience. *Ann Tourism Res* 38:499–519
112. Riddington G, McArthur D, Harrison T, Gibson H (2010) Assessing the economic impact of wind farms on tourism in Scotland: GIS, surveys and policy outcomes. *Int J Tourism Res* 12:237–252
113. Lilley MB, Firestone J, Kempton W (2010) The effect of wind power installations on coastal tourism. *Energies* 3:1–22
114. Gross C (2007) Community perspectives of wind energy in Australia. The application of a justice and community fairness framework to increase social acceptance. *Energy Policy* 35:2727–2736
115. Mumford J, Gray D (2010) Consumer engagement in alternative energy – can the regulators and suppliers be trusted? *Energy Policy* 38:2664–2671
116. Frey BS, Oberholzer-Gee F (1997) The cost of price-incentives: an empirical analysis of motivation crowding out. *Am Econ Rev* 87:746–755
117. Aitken M (2010) Wind power and community benefits: challenges and opportunities. *Energy Policy* 38:6066–6075
118. Rogers EM (1995) Diffusion of innovations, 4th edn. The Free Press, New York
119. Breukers S, Wolsink M (2007) Wind power in changing institutional landscapes: an international comparison. *Energy Policy* 35:2737–2750
120. Zarnikau J (2003) Consumer demand for ‘green’ power and energy efficiency. *Energy Policy* 31:1661–1672
121. Wüstenhagen R, Bilharz M (2006) Green energy market development in Germany: effective public policy and emerging customer demand. *Energy Policy* 34:1681–1696
122. MacKay DJC (2008) Sustainable energy – without the hot air. UIT, Cambridge
123. Bolinger MA (2005) Making European-style community wind power development work in the US. *Ren Sust Energy Rev* 8:556–575
124. Han JY, Mol APJ, Lu YL, Zhang L (2009) Onshore wind power development in China: challenges behind a successful story. *Energy Policy* 37:2941–2951
125. Bergek A, Jacobsson S, Sanden BA (2008) ‘Legitimation’ and ‘development of positive externalities’: two key processes in the formation phase of technological innovation systems. *Technol Anal Strateg* 20:575–592
126. Bird L, Bolinger M, Gagliano T, Wiser R, Brown M, Parsons B (2005) Policies and market factors driving wind power development in the United States. *Energy Policy* 33:1397–1407
127. Lewis JI, Wiser RH (2007) Fostering a renewable energy technology industry: an international comparison of wind industry support mechanisms. *Energy Policy* 35:1844–1857
128. Bansal P, Roth K (2000) Why companies go green: a model of ecological responsiveness. *Acad Manage J* 43:717–736
129. Kaldellis JK, Kavadias KA, Paliatsos AG (2004) Environmental impacts of wind energy applications: “Myth or reality?”. *Fresen Environ Bull* 13:413–423
130. Bishop I (2002) Determination of thresholds of visual impact: the case of wind turbines. *Environ Plann B* 29:707–718
131. Kaldellis JK (2006) Evaluation of Greek wind parks visual impact “public attitude and experts’ opinion”. *Fresen Environ Bull* 15:1419–1426
132. Brittan GG (2001) Wind, energy, landscape: reconciling nature and technology. *Philos Geogr* 4:169–184
133. Mercer D (2003) The great Australian wind rush and the devaluation of landscape amenity. *Austral Geogr* 34:91–121
134. Johansson M, Laike T (2007) Intention to respond to local wind turbines: the role of attitudes and visual perception. *Wind Energy* 10:435–451
135. Zografos C, Martinez-Allier J (2009) The politics of landscape value: a case study of wind farm conflict in rural Catalonia. *Environ Plann A* 41:1726–1744

136. Bishop ID, Miller DR (2007) Visual assessment of off-shore wind turbines: the influence of distance, contrast, movement and social variables. *Renew Energ* 32:814–831
137. Ladenburg J, Dubgaard A (2009) Preferences of coastal zone user groups regarding the siting of offshore wind farms. *Ocean Coast Manage* 52:233–242
138. Haggett C (2008) Over the sea and far away? A consideration of the planning, politics and public perception of offshore wind farms. *J Environ Pol Plann* 10:289–306
139. Ladenburg J (2009) Visual impact assessment of offshore wind farms and prior experience. *Appl Energ* 86:380–387
140. Gonzalez MI, Estevez B (2005) Participation, communication and negotiation in environmental conflicts: offshore wind energy in the Trafalgar Sea area. *Arbor-Ciencia Pensamiento y Cultura* 181:377–392
141. Phadke R (2010) Steel forests or smoke stacks: the politics of visualisation in the Cape Wind controversy. *Environ Polit* 19:1–20
142. Wolsink M (2010) Near-shore wind power – protected seascapes, environmentalists' attitudes, and the technocratic planning perspective. *Land Use Policy* 27:195–203
143. Bell S (1999) *Landscape: pattern, perception and process*. Routledge, New York
144. Lothian A (1999) Landscape and the philosophy of aesthetics: is landscape quality inherent in the landscape or in the eye of the beholder? *Landscape Urban Plan* 44:177–198
145. Parsons R, Daniel TC (2002) Good looking: in defense of scenic landscape aesthetics. *Landscape Urban Plan* 60:43–56
146. Jay S (2010) Planners to the rescue: spatial planning facilitating the development of offshore wind energy. *Mar Pollut Bull* 60:493–499
147. Gray T, Haggett C, Bell D (2005) Offshore windfarms and commercial fisheries in the UK: a study in stakeholder consultation. *Ethics Place Environ* 8:127–140
148. Michler-Cieluch T, Krause G, Buck BH (2009) Reflections on integrating operation and maintenance activities of offshore wind farms and mariculture. *Ocean Coast Manage* 52:57–68
149. Sovacool BK (2009) Contextualizing avian mortality: a preliminary appraisal of bird and bat fatalities from wind, fossil-fuel and nuclear electricity. *Energ Policy* 37:2241–2248
150. Solli J (2010) Where the eagles dare? Enacting resistance to wind farms through hybrid collectives. *Environ Polit* 19:45–60
151. Nadaï A, Labussière O (2010) Birds, wind and the making of wind power landscapes in Aude, Southern France. *Landscape Res* 35:209–233
152. Verkuijlen E, Westra CA (1984) Shadow hindrance by wind turbines. In: European wind energy conference, Hamburg, pp 356–361
153. Wolsink M, Sprengers M (1993) Windturbine noise: a new environmental threat? In: Vallet M (ed) *Noise as a public health problem*, vol 2. INRETS, Bron, pp 235–238
154. Pedersen E, Persson-Waye K (2007) Wind turbine noise, annoyance and self-reported health and well-being in different living environments. *Occup Environ Med* 2007:480–486
155. Pedersen E, Larsman P (2008) The impact of visual factors on noise annoyance among people living in the vicinity of wind turbines. *J Environ Psychol* 28:379–389
156. Dalby S, Mackenzie F (1997) Re-conceptualising local community: environment, identity and threat. *Area* 29: 99–108
157. Walker G, Devine-Wright P, Hunter S, High H, Evans B (2010) Trust and community: exploring the meanings, contexts and dynamics of community renewable energy. *Energ Policy* 38:2655–2663
158. Higgs G, Berry R, Kidner D, Langford M (2008) Using IT approaches to promote public participation in renewable energy planning: prospects and challenges. *Land Use Policy* 25:596–607
159. Bishop ID, Stock C (2010) Using collaborative virtual environments to plan wind energy installations. *Renew Energ* 35:2348–2355
160. Chicco G, Mancarella P (2009) Distributed multi-generation: a comprehensive view. *Ren Sust Energ Rev* 13:535–551
161. Gardner J, Car-Cornish SG, Ashworth PN (2008) Exploring the acceptance of a domestic distributed energy market in Australia. *Australasian J Environ Man* 15:93–103
162. Charles D (2009) Renewables test IQ of the grid. *Science* 324:172–175
163. Stadler I (2008) Power grid balancing of energy systems with high renewable energy penetration by demand response. *Utilit Policy* 16:90–98
164. Green RC, Wang L, Alam M (2011) The impact of plug-in hybrid electric vehicles on distributed networks: a review and outlook. *Ren Sust Energ Rev* 15:544–553
165. Marris E (2008) Upgrading the grid. *Nature* 454:570–573

Books and Reviews

- Agterbosch S (2006) Empowering wind power. On social and institutional conditions affecting the performance of entrepreneurs in the wind power supply market in the Netherlands. *Netherland Geographical Studies* 351, KNAG, Utrecht
- Breukers SC (2007) Institutional capacity building for wind power, a geographical comparison. Vossius Pers/Amsterdam University Press, Amsterdam, <http://dare.uva.nl/record/209600> [Open access]
- Devine-Wright P (ed) (2011) *Renewable energy and the public: from NIMBY to participation*. Earthscan, London
- Ellis G, Cowell R, Warren C, Strachan P, Szarka J, Hadwin R, Miner P, Wolsink M, Nadai A (2009) Interface. *Wind power: is there a "planning problem"?* *Plann Theory Pract* 10:521–547
- Jay SA (2008) *At the margins of planning. Offshore wind farms in the United Kingdom*. Ashgate, Aldershot
- Kahn J (2004) *Local politics of renewable energy. Project planning, siting conflicts and citizen participation*. Lund University, PhD thesis
- MacKay DJC (2008). Sustainable energy – without the hot air. www.withouthotair.com [Open access]

- Nadaï A, van der Horst D (eds) (2010) Wind power planning, landscapes and publics. *Land Use Policy* 27(Special Issue):181–254
- Scheer H (2002) The solar economy: renewable energy for a sustainable global future. Earthscan, London
- Sovacool B (2008) The dirty energy dilemma: what's blocking clean power in the United States? Springer, Dordrecht
- Szarka J (2007) Wind power in Europe: politics, business and society. Macmillan Palgrave, Basingstoke
- Wüstenhagen R, Wolsink M, Bürer MJ (eds) (2007) Social acceptance of renewable energy innovation. *Energy Policy* 35(Special Issue):2683–2810

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit

MICHAEL MILLIGAN¹, LORI BIRD¹, SUZANNE TEGEN¹, KEVIN PORTER², SARI FINK², BRENDAN KIRBY¹

¹National Renewable Energy Laboratory, Golden, CO, USA

²Exeter Associates, Columbia, Maryland, USA

Article Outline

Glossary
 Definition of Subject
 Introduction
 Wind Energy in Wholesale Markets
 Wind Energy and Energy Markets in the United States
 Wind Energy and Energy Markets in Europe
 Wind Energy and Capacity Markets
 Wind Energy Tax Credits
 Green Power Markets
 Capacity Credit
 Wind Integration Costs
 Cost Trajectory for Wind Energy Technology
 Wind Energy and Economic Development
 Future Directions
 Bibliography

Glossary

Ancillary services Additional services that are required to manage the delivery of energy. Examples include reactive power support, balancing and regulation, and operating reserves.

Balancing area The collection of generation, transmission, and loads within a metered boundary where a power system operator maintains the load-resource balance.

Capacity markets Market-based method of procuring sufficient generation and demand response to assure that future load and reserve requirements will be met.

Day-ahead market Hourly wholesale electricity market that is cleared the day before the operating day.

Real-time market Hourly or sub-hourly wholesale market that is cleared shortly before the operating interval.

Reserves (operating) Additional generating capacity that is available in real time or near real time if needed to compensate for unanticipated generator failures (contingency reserves) or unanticipated changes in load or wind generation (load-following reserve).

Reserves (planning) Generating capacity that is built in excess of predicted load in a future time period. This excess reserve margin is calculated to account for unexpected failures of power plants, ensuring an adequate supply of electricity at all times.

Self-schedule Decision by a generator operator to operate at a specific time regardless of price.

Unit commitment The process of determining the optimal mix of generators that must be started or otherwise made available one or more days in advance of need.

Wind energy penetration Percentage of wind relative to the rest of the system. This is typically calculated using energy on an annual basis; the percentage of total energy demand for a year that wind energy provides.

Wholesale electricity markets Wholesale bulk power market that facilitates the exchange of energy among buyers and sellers.

Definition of Subject

With increasing concern regarding climate change and emissions from conventional power plants, new markets have developed that allow for the trading of the clean energy attributes of wind and other renewable energy sources. We discuss these green power markets and their role in demonstrating compliance with renewable energy

standards. Wind energy can also be bought and sold in wholesale energy markets, exchanges where bulk energy is bought and sold. In some areas, capacity markets exist that provide payment for installed capacity that is available when needed. Capacity markets are generally based on the capacity contribution that a power plant or responsive load can make toward planning reserves, which is called the capacity credit or capacity value. The general motivation behind a capacity market is to ensure sufficient revenue for the generator and sufficient installed capacity.

In some countries, government incentives exist that encourage the development and use of wind energy (or other renewable energy sources). These may include the following:

1. Renewable energy standards, sometimes called renewable portfolio standards: These standards specify some percentage of all energy or capacity that must be provided by renewable energy resources.
2. Wind energy tax credits: reduction in tax liability to encourage development and use of wind energy. The tax credits can be based on the amount of energy production from the wind plant, or can be based on investment or other economic attribute.
3. Green power markets: a market where the “green” attributes of renewable energy can be traded. The objectives of these markets can be to help in satisfying compliance rules for renewable energy standards, or could be alternatives to policies that limit the emissions (including carbon) from conventional power plants.

Because of its variable and uncertain characteristic, wind energy can increase overall operating reserves that are needed to manage uncertainty. The costs associated with this increase in reserves and changes in the way the system is dispatched are called wind integration costs. Specific methods to measure this cost have evolved, and today there is some controversy surrounding the topic and its precise definition. Integration impacts and costs are not unique to wind. Nuclear power, for example, typically requires larger contingency reserves and increases cycling of other base-load generation. All investments in the power system impact system operations and costs.

With the recent increase in actual and planned wind-generated electricity, there has been an accompanying

increase in estimating the economic development impacts of wind. These impacts result from expenditure on the construction and operation of wind turbines or plants. Much of the focus of economic development impact studies is on jobs related to the wind plant, and the change in spending that occurs as a result.

All of these issues are potentially important as decision makers evaluate policies that may influence the development of wind energy.

Introduction

Wind energy has seen significant growth in the past several years. As a result, it has become more visible to power system planners and operators, electricity wholesale market designers, and to policy makers. During the same period, there have been substantial changes in the way that electricity systems are operated. The single major change has been the movement toward large coordinated wholesale electricity markets. The confluence of these two development trends has resulted in a substantial body of analysis and development of market mechanisms that are intended to promote economically efficient exchange of electricity, taking into account the characteristics of the many generating technologies that are used today, and in the near future.

Wind energy is integrated, bought, and sold using a variety of markets and mechanisms. In this entry we examine several key wind energy issues related to economic development, wholesale electricity markets, subsidies, and capacity value. Because wind energy is integrated into a large complex power system, wind issues often must be analyzed in that context. Electricity markets, for example, apply to all buyers and sellers in a particular region. Because there are other generation technologies in use on the grid, markets must be designed and operated in such a way that the characteristics of all technologies can be recognized that result in economically efficient power system operation, achieving appropriate reliability targets.

This entry examines the characteristics of the wholesale electricity markets in which wind energy operates, discusses the capacity value and implication for the acquisition of other generation to ensure generation adequacy, the role of tax credits, integration cost impacts, and the economic development impacts

of wind energy development. The entry concludes with a brief assessment of future directions.

Wind Energy in Wholesale Markets

The way wind energy participates in wholesale energy markets can vary significantly. In this section we describe some of the key approaches used in the United States and in Europe. We note that wholesale electricity markets in many parts of the world are currently undergoing changes that may alter some of the discussion contained in this entry. Market characteristics discussed herein were in place at the time of this writing. Common elements in international comparisons include the basic operation of wholesale electricity markets. These markets help manage the bulk power system (combination of large generators and high-voltage transmission) balance, which requires total generation to equal total load at all times. When operated on short time steps, such as 5 or 10 min, these markets can provide the power system operator with access to generation on a sub-hourly basis, subject to generator economics. If electricity demand is increasing, generating units that have offered to increment their output can be chosen to do so, based on generation price bids. In some regions (e.g., much of the Western portion of the United States) energy markets are based more on bilateral contracts (private contracts between any two parties on their own terms for their own reasons), and operate on hourly time steps across balancing areas that are sometimes only weakly interconnected. Hourly scheduling and dispatch restricts access to the physically available generation, which may be capable of responding but cannot be accessed by the power system operator. Generators that may be physically capable of responding, but that are not economic, will not be called upon in favor of other less expensive generation.

Well-functioning markets provide for economically efficient exchanges between buyers and sellers, with transparent information available for all parties. Economic efficiency is attained when markets provide the goods and services that are most desired by consumers and society, at lowest possible cost. Markets can also help efficiently integrate wind energy [1].

In some market areas there are ancillary products that are also exchanged. These can include

capacity markets, and various types of reserve products, such as grid-synchronized fast response reserves, nonsynchronized slow response, contingency reserves, and regulating reserves. The terminology in Europe differs from that in the United States, but the general properties are similar.

In addition to providing an institutional mechanism to tap into existing energy supplies or ancillary services, markets provide signals to investors about relative scarcity and value. In perfectly competitive markets, these prices and all other information are fully transparent, and investment occurs in an economically efficient manner. If wholesale electricity markets function in this way, prices would induce developers to ensure that appropriate generating resources and transmission are available at future dates when needed. Typically, the relatively small number of sellers results in varying degrees of imperfect competition in wholesale electricity markets, resulting in at least some loss of economic efficiency. When energy markets are unable to provide sufficient generation capability to provide needed services such as ramping or minimum generation levels that may be needed to complement wind energy at high penetrations, other supplementary markets may need to be developed.

Wind Energy and Energy Markets in the United States

There are seven regional transmission organizations (RTOs) and independent system operators (ISOs) managing parts of the electric grid in the United States – PJM Interconnection (PJM), Midwest ISO, New York ISO (NYISO), ISO New England (ISO-NE), California ISO (CAISO), Electric Reliability Council of Texas (ERCOT), and the Southwest Power Pool (SPP). These entities operate the grid for approximately 65% of total US electricity demand [2]. PJM, the NYISO, the CAISO, the Midwest ISO, and ISO-NE, all operate wholesale energy markets, based on location-specific prices known as locational marginal prices (LMPs) that are for both day-ahead and real-time markets, with generation dispatched in real time at 5-min intervals. SPP and ERCOT are currently in the process of implementing LMP-based markets with plans to have fully functioning wholesale markets in place within the next few years. Three of the RTOs/ISOs (PJM, NYISO,

ISO-NE) also administer capacity markets or capacity auctions. Another ISO, the Midwest ISO, conducts a residual capacity auction where load-serving entities (LSEs) can make up shortfalls. The vast majority of capacity however, is obtained via bilateral contracts and the Midwest ISO's auctions are not used to any significant extent.

Other regions, particularly the West and Southeast, are not in RTOs nor do they have RTO-administered energy or capacity markets. Instead, they have hourly bilateral markets and utilities are responsible for meeting their own capacity obligations. Wind energy generation is variable and uncertain, and therefore not dispatchable in the traditional sense like conventional generation has been operated. In non-RTO markets, any variations in output from day-ahead schedules and real-time output are subject to the energy imbalance provisions of Order 890 that was issued by the Federal Energy Regulatory Commission (FERC) in 2007. Imbalances are differences between the scheduled and actual flows from generators or to loads. Imbalances of less than or equal to 1.5% of scheduled energy, or up to 2 MW (whichever is greater), would be netted monthly and settled at the power system operator's incremental or decremental costs. Imbalances of between 1.5% and 7.5% of schedule energy, or between 2 and 10 MW (whichever is larger), would be settled at 90% of decremental costs and 110% of incremental costs. ISO and RTOs have different methods of handling energy imbalances from wind power – these are outlined in [Table 1](#).

As wind penetration increases, some power system operators are taking steps to more actively incorporate wind in their dispatch. The NYISO, for instance, has integrated wind plants into its economic dispatch system and treats wind resources like all other generation. In the NYISO markets, generators indicate their willingness to redispatch through their economic offers. Wind facilities, like all generators, are required to submit economic price curves into the real-time market, due 75 min before each operating hour (optional for day-ahead market). The wind energy price curves consist of price–quantity pairs that indicate the LMPs at which each wind facility is willing to operate. The price curve can include up to 11 price–quantity pairs, which can be changed with each bid, and can also include negative prices. Wind facilities that receive a production tax credit (PTC) may be willing

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 1 Energy imbalance settlement

Entity	Imbalance settlement
PJM	For wind resources that are self-scheduled, operating reserve deviation charges apply on differential between day-ahead and real-time levels exceeding a dead band; differentials less than 5% or 5 MW incur no deviation charges. Wind resources that are not self-scheduled are assessed deviation charges based on how closely they follow PJM dispatch signals.
NYISO	If scheduling day-ahead, buy out shortfalls at real-time LMPs. Up to 3,300 MW exempt from under-generation penalties during unconstrained operations.
ISO-NE	If scheduling day-ahead, deviations are settled at real-time LMP.
Midwest ISO	Settled at real-time price with no deviation penalties.
SPP	Wind is not subject to uninstructed deviation charges.
CAISO	If participating in wind forecasting, then hourly deviations are accumulated for monthly average of energy imbalances and settled at a monthly weighted market-clearing price and. If not participating in wind forecasting, then subject to 10-min imbalance energy charges.
ERCOT	Any imbalances are settled at zonal market clearing prices. No take back of payment for over generation if within $\pm 50\%$ of scheduled capacity. Wind-only scheduling entities are exempt from deviation charges.

Source: Utility Wind Interest Group. *Wind power and electricity markets*. August 2009. <http://www.uwig.org/WindinMarketsTableSept09.pdf>

to continue providing energy even when wholesale prices drop below zero because payments such as the PTC are received by wind plants any time they produce energy; thus, the wind plant can earn revenue when price is negative as long as price exceeds the negative of the PTC. The NYISO's Security Constrained Economic Dispatch software (SCED) recognizes that wind facilities cannot increase production on demand, and therefore, the offers from wind plants are taken as the prices below

which each wind facility no longer wishes to produce energy. The NYISO has also implemented a centralized wind forecasting system. All wind plants must meet the technical requirements for participating in wind forecasts and are required to provide NYISO with meteorological data. The wind forecasts are used to determine day-ahead generator commitments and are integrated into the SCED. The NYISO evaluates all generator offers, the wind forecast, and the load forecast to produce a least-cost dispatch solution every 5 min, then sends automatic dispatch signals to all generators.

Wind resources in PJM, ISO-NE, and the Midwest ISO can submit bid curves into the day-ahead and real-time markets.

1. In PJM, if a wind facility has cleared as a capacity resource through the forward capacity market the facility must offer into the day-ahead market. In early 2009, PJM reprogrammed its dispatch software to allow for negative price offers and implemented a centralized wind forecasting system with plans to integrate the forecasts into their market operations.
2. In the Midwest ISO, if a wind facility has a capacity obligation it must offer or self-schedule into the day-ahead market. The Midwest ISO has a centralized wind forecasting system that is used for reliability, outage coordination, and analysis, but not for market operations.
3. In ISO-NE wind facilities that have cleared as a capacity resource in the forward capacity market must bid or self-schedule into the real-time market. ISO-NE does not currently utilize wind forecasting.

Noncapacity designated wind facilities participating only in the real-time markets receive real-time LMP for the energy provided. Wind plants that do not qualify or do not want to receive capacity payments do not participate in the day-ahead market. Instead, they accept payment based on the LMP clearing price for each market interval.

CAISO implemented day-ahead and real-time markets in 2009 but wind facilities are still considered price takers, which can if they wish, sell energy into the real-time markets at real-time LMP. SPP and ERCOT are in the process of creating wholesale LMP-based markets but at current, energy transactions are conducted bilaterally by the relevant entities. CAISO and ERCOT both

have centralized wind forecasting systems while SPP requires balancing authorities or wind generators to provide forecasts. How energy from wind facilities is treated in CAISO, SPP, and ERCOT may change in the future as their wholesale electricity markets evolve.

Because wind energy has a near-zero marginal cost, it is nearly always economically efficient to accommodate the wind energy whenever it is available, backing down other generation sources that require fuel. We note that exceptions may exist in cases that involve a constrained transmission interface that causes wind to be curtailed. Curtailment need not be for the full amount of wind generation during the critical time period. Energy prices in the market are often set by the cost to generate one additional MWh of energy, and can vary by location depending on grid congestion, load, wind, and mix of other generation. At high wind energy penetrations, the energy price tends to be driven down (as it would be with the addition of any low-marginal-cost energy provider such as nuclear), reducing the output, and therefore the annual capacity factors, of conventional generators. This can reduce the realized revenue stream, and depending on the level of load and wind, energy prices can be very low or even negative in extreme cases. It is unclear whether the resulting price volatility will induce economically efficient levels of different types of generation in the long run. This is a general concern with electricity pricing issues and not unique to wind.

Wind Energy and Energy Markets in Europe

Just as in the North America, there is a great deal of diversity in the structural details of European electricity markets with some reasonably consistent characteristics. Day-ahead markets are working in most countries. Intra-day hourly markets exist in some countries. Fifteen-minute balancing markets exist in a few countries. While half of the US load is served in regions with 5-min markets, no 5-min markets exist in Europe. The European Network of Transmission System Operators for Electricity (ENTSO-E) is working to harmonize markets throughout Europe with long-term, day-ahead, intra-day and balancing markets. The ENTSO-E took over all operational tasks of the six existing TSO associations in Europe in July 2009 and now encompasses 42 TSOs from 34 countries [3].

Comparing North American and European power systems, or discussing them in the same paper, is complicated by differences in the terminology used in each location. “Regulation” in North America, for example, refers to the minute-to-minute balancing ancillary service that is procured through hourly markets in ISO/RTO regions and obtained from dedicated generation capacity in all regions. In Europe, Transmission System Operators (TSOs) operate hourly energy markets that are often referred to as “regulation” balancing markets. North American “regulation” is obtained from secondary reserves in Europe. Secondary reserves (centralized power system operator control of generation through automatic generator control) are typically obtained through mandatory requirements in Europe rather than through day-ahead hourly regulation markets as is done in much of North America. Wind “regulation” requirements and costs in European studies typically refer to intra-day energy scheduling costs. Wind advocates in Europe recognize the benefits of shorter gate closure times, faster market clearing, and larger geographic market scope, and are encouraging market harmonization [4].

The Nordic countries, Germany, France, and the Netherlands provide an example of typical European electricity markets structure. As in North America, a series of markets clear successively as real time is approached [5]:

1. Bilateral over the counter electricity trading: trading takes place between individuals outside of the public markets.
2. Day-ahead hourly market (spot market), which close between 10 am and noon of the day-ahead.
3. Intra-day hourly market is another hourly market that operates within the day. Markets typically clear 2 h ahead (75 min ahead in Germany). Some countries have 2-h intervals rather than hourly.
4. Intra-hour market(s). This is in fact two linked markets operating at 15-min intervals. The TSO acts as the sole buyer in the first and the sole seller in the second and facilitates the sub-hourly transactions.
 - (a) “Regulating power market” (RPM). A real-time market for the TSO to purchase energy to counter imbalances from the day-ahead schedules. Both generators and loads can be

sold into the market. Typically, all suppliers get the market clearing price but in Denmark each supplier gets its bid price. Settlements are typically determined after the fact.

- (b) “Balancing market.” A real-time market that is linked to the RPM. The TSO is the sole supplier in this market and both generators and loads are price takers to settle imbalances from day-ahead schedules.

In Denmark, the RMP is subdivided according to the different service response times (secondary and tertiary reserves). In the Netherlands, TenneT (the TSO) contracts annually for reserves and obtains 15 min balancing energy (regulating power) through least-cost bids. Spain assigns primary reserve requirements to all generators and can adjust the amount annually. Secondary reserves are obtained day-ahead through a bidding process. Tertiary reserves are obtained through 15 min energy scheduling. In Belgium, the Access Responsible Parties are responsible for balancing portfolios in the day-ahead and hourly time frame. The TSO is responsible for balancing the aggregate system with secondary and tertiary reserves [6].

Cross-border trading typically has, depending on applied methods, less scheduling flexibility than that provided by within-country markets, effectively limiting wind and load aggregation benefits. An exception is in the Nordic market in Northern Europe where there is a common regulating market between four countries where the cheapest bids in the four countries is used if there is physical room on the transmission. This means, for example, that wind power changes within 15 min in Denmark can be balanced in Finland if the cheapest bid is in Finland and there is transmission capacity available through Sweden.

There are also differences in how wind is treated within each national market. Wind is outside the markets in France, which uses a feed-in tariff. In Germany, the four TSOs are required to convert variable wind into flat profiles equal to the long-term average wind generation. Wind is balanced as a separate balancing group. The costs for balancing wind are socialized in Germany. Wind plants in the Netherlands are always combined with other loads and generation as a part of a balanced portfolio from a Program Responsible Party.

Wind Energy and Capacity Markets

It is common practice in the electric power industry to maintain additional generating capacity (or to be able to call upon customers to reduce demand) to ensure that demand for electricity is met in case demand is higher than expected, or a generating or transmission project is unexpectedly out of service. Such additional generating capacity or demand reduction capability is collectively known as reserve capacity. How much more reserve capacity is maintained varies by power system operator, but 10–15% is a common rule of thumb for planning reserves, whereas lesser amounts are used in the operation time frame.

Reserve requirements are managed differently throughout the United States, whereas Europe generally does not have capacity markets. In some cases, a state regulatory entity or regional reliability entity sets the reserve requirement and utilities are responsible for meeting that requirement through their own generating or demand resources, or through bilateral contracts with other utilities, generators, power marketers, or demand response providers. In the north-eastern United States (but not including New York), annual capacity auctions are deployed to solicit needed capacity 3 years in advance of identified need, with incremental auctions if additional capacity is needed. The following are some examples of how reserve capacity is acquired:

1. California ISO

California is an example of a jurisdiction that requires load-serving entities (LSEs) to self-supply capacity through with their own generation ownership of facilities or through bilateral contracts. The California Public Utilities Commission (CPUC), along with the California ISO, sets an annual reserve requirement and also a forward resource adequacy requirement for the next operating year. LSEs must demonstrate they have sufficient capacity reserves 1 year in advance. Historical wind generation during the peak period is used to determine wind's contribution to resource adequacy.

2. New York ISO

The New York ISO (NYISO) is an example of a jurisdiction that provides a capacity market which acts as a backstop procurement mechanism in addition to generation ownership by LSEs and bilateral

contracting. The majority of LSEs opt to procure capacity outside the centralized ISO market. The NYISO capacity requirement also includes a locational element, and LSEs are required to make up capacity shortfalls for each zone through purchases in the twice-yearly backstop capacity auctions. New York uses historical wind production data and reliability-based techniques to assess wind's contribution to resource adequacy.

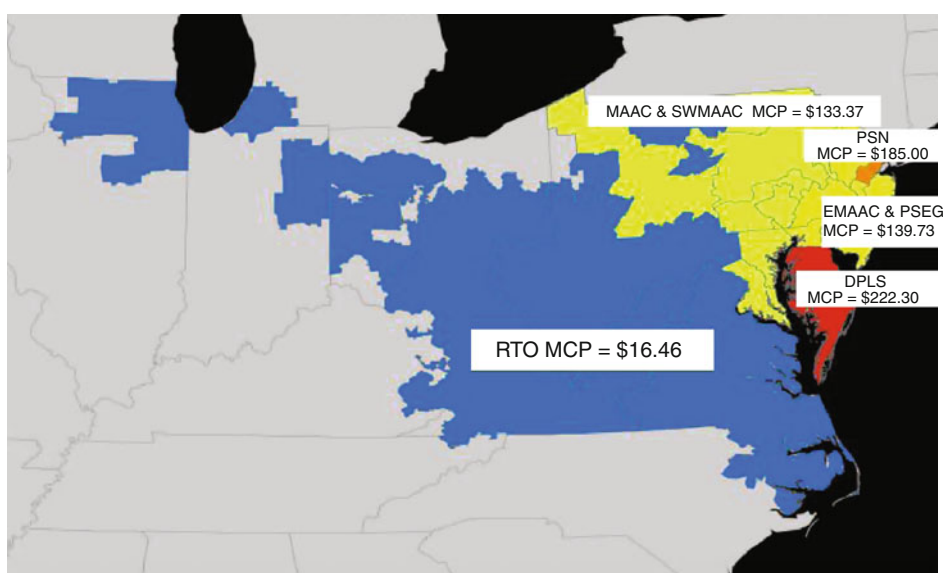
3. PJM Interconnection and ISO New England

PJM and ISO New England (ISO-NE) administer energy markets and have both implemented 3-year forward capacity requirements through centralized forward markets. LSEs are required to obtain adequate capacity to meet reserve requirements 3 years in advance through the forward capacity markets. LSEs can opt to self-supply capacity through ownership or bilateral contracts for capacity via a special arrangement, but must commit to doing so for a set period. The vast majority of LSEs opt to obtain capacity through the forward markets. Under the PJM's Reliability Pricing Model (RPM) capacity market and the ISO-NE's Forward Capacity Market (FCM), auctions are held 3 years in advance of the delivery year, with supplemental annual incremental auctions where LSEs can purchase capacity to make up shortfalls or sell excess reserves. Wind facilities can bid their capacity into the PJM and ISO-NE capacity auctions and receive the per-MW market clearing price. Prices for capacity can vary greatly from region to region and from auction to auction. [Table 2](#) shows the capacity prices for the four major zones for PJMs base-year RPM auctions that have been held as of May 2010.

PJM partitions the region into zones based on individual utility territories (see [Fig. 1](#) for the 2012/2013 zonal results) and calculates the capacity reserve requirements by zone. The zonal capacity price differentials are mainly due to transmission capacity constraints. Depending on where the constraints are located, the zonal groupings can vary from auction to auction. For example, in the 2012/2013 auction, DPLS was a constrained utility area and received an individual zonal clearing price, while in other auctions the price has been the same as in EMAAC. The majority of the low-cost generation capacity is located within RTO MCP

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 2 Capacity prices for PJMs base-year RPM auctions (as of May 2010)

Delivery year	RTO	MAAC	SWMAAC	EMAAC
June 2007–May 2008	\$40.80/MW-day	\$40.80/MW-day	\$188.54/MW-day	\$197.67/MW-day
June 2008–May 2009	\$111.92	\$111.92	\$148.80	\$210.11
June 2009–May 2010	\$102.04	\$191.32	\$191.32	\$237.33
June 2010–May 2011	\$174.29	\$174.29	\$174.29	\$174.29
June 2011–May 2012	\$110.00	\$110.00	\$110.00	\$110.00
June 2012–May 2013	\$16.46	\$133.37	\$133.37	\$139.73
June 2013–May 2014	\$27.73	\$226.15	\$226.15	\$245.00



Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Figure 1

PJM RPM auction zones and clearing prices for 2012/2013 auction (Source: PJM 2012/2013 BRA clearing price map, May 27, 2009)

zone, while the majority of PJM load is in the east. PJM reported that 323 MW of wind capacity cleared in the 2012/2013 auction, representing 13% of PJM's total nameplate wind capacity. PJM sets a wind capacity factor limit at 13% of nameplate unless a wind facility can provide historical data proving a higher capacity factor is warranted. PJM's 13% capacity factor limit for wind plants that do not yet have historical experience is based

on PJM analysis of PJM's past experience with wind at that time. Therefore, wind facilities can only bid 13% of nameplate capacity into the RPM markets.

The ISO-NE FCM contains two zones, Maine and Rest of Pool, and has held three auctions as of the end of 2009. These auctions all yielded the same clearing price for both zones. The clearing prices and the amount of

wind capacity that cleared in each FCM auction are outlined in [Table 3](#). We note that ISO-NE reports clearing prices in kW-months (the price of a kW of generation or demand response capacity available for a month). The table therefore contains approximate conversions to MW-days (the price of a MW of generation or demand response capacity available for a day).

Wind Energy Tax Credits

Several tax incentives have served to boost utility-scale wind energy development in the United States. This discussion focuses on wind energy but the tax credits detailed here were/are also available for other eligible resources. Investment tax credits (ITCs) were the first federally enacted incentive utilized for promoting wind energy development. The ITC for wind was introduced in the Energy Tax Act of 1978, and provided a 10% federal tax credit on new investment in wind generation technologies. ITCs are tax credits equal to a percentage of the value of qualifying tangible, depreciable property. Such credits directly offset the investing taxpayer's federal income tax liability and are claimed on the tax return for the year a wind turbine and related equipment is first placed in service. In 1981, Congress shortened the depreciation period for wind technologies from 18 years to 5 years (which are still in effect today). Combined with state incentives and favorable power purchase rates, the ITC and the accelerated depreciation helped kick-start the modern wind industry, particularly in California. Total installed wind capacity increased from 10 MW in 1981 to 1,039 MW in 1985. However, the ITC expired in 1985, and a sharp decrease in fossil fuel prices in 1986 led to

a drop in power purchase prices to levels that were uneconomic for new wind projects. As a result, and after installing over 1,000 MW in 5 years, US wind capacity installations plummeted, with only 500 MW of wind capacity installed between 1987 and 1998.

The wind industry received a new tax incentive when Congress approved a PTC as part of the Energy Policy Act of 1992. The PTC is tied to energy production and provides a per-kWh tax credit for energy produced from a qualifying renewable facility. The PTC is currently (2010) set at 2.1¢/kWh over the first 10 years of operation for wind energy and is adjusted annually for inflation. The PTC has been renewed several times and was allowed to lapse a few times. The latest renewal was in 2009 as part of the American Recovery and Reinvestment Act (ARRA) for wind projects that go into service by the end of 2012. [Table 4](#) shows the history and duration of the PTC.

The ARRA also extended a 50% depreciation bonus that was first made available in 2008, which allows a project developer to deduct half the cost of a project when it is first placed in service. This depreciation bonus is a limited-time offer available only for projects completed and placed in-service by the end of 2009.

The availability (or lack thereof) of tax credits has affected US wind project development over the last 3 decades. Wind energy development activity seems to correlate with the availability of the ITC in the early 1980s and the PTC, which lapsed in 2002 and 2004 (see [Fig. 2](#)).

In addition to extending the PTC through 2012, the ARRA also created an option to choose an Investment Tax Credit (ITC) or a cash grant instead of the PTC. PTC-qualified wind facilities can opt to receive a 30% ITC in lieu of the PTC but the depreciable basis of property must be reduced by half of the ITC amount. The cash grant option is available for projects placed in service in 2009 or 2010 or that started construction by the end of 2010 and are in service by the end of 2012. The grant can be up to 30% of the total cost of the project and applications must be submitted by October 1, 2011. The grant is excluded from gross income and, as with the ITC, the depreciable basis of the property must be reduced by half the amount. On September 1, 2009, the Treasury and the DOE announced the first round of grants in lieu of the PTC/ITC through the ARRA was being distributed to two solar projects and

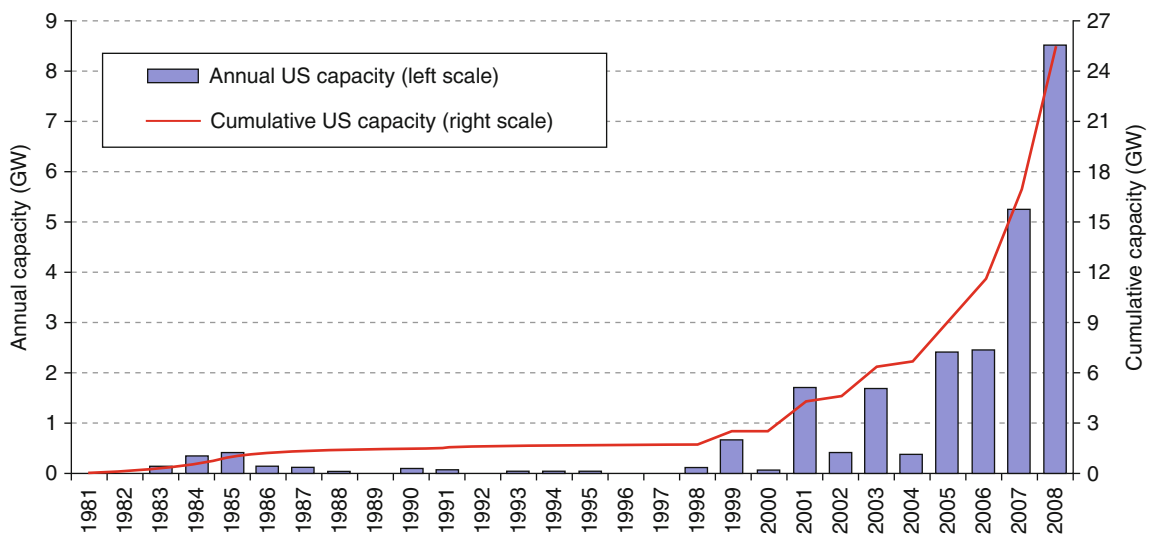
Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 3 Clearing prices and wind capacity in FCM auctions

Delivery year	Capacity clearing price	Wind capacity (MW)
June 2010–May 2011	\$148/MW-day ^a	10
June 2011–May 2012	\$118	81
June 2012–May 2013	\$97	93

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 4 PTC history and duration

Legislation	Date enacted	PTC eligibility window (for wind)	PTC lapse duration	Effective duration of PTC window (considering lapses)
Energy Policy Act of 1992	01/24/92	1994–June 1999	n/a	80 months
Ticket to Work and Work Incentive Improvement Act of 1999	12/19/99	July 1999–2001	6 months	24 months
Job Creation and Worker Assistance Act	03/09/02	2002–2003	2 months	22 months
The Working Families Tax Relief Act	10/04/04	2004–2005	9 months	15 months
Energy Policy Act of 2005	08/08/05	2006–2007	None	24 months
Tax Relief and Health Care Act of 2006	12/20/06	2008	None	12 months
Emergency Economic Stabilization Act of 2008	10/03/08	2009	None	12 months
American Recovery & Reinvestment Act of 2009	02/13/09	2010–2012	None	36 months

Source for original table: [2]. Modified to include EESA 2008 and ARRA 2009.

**Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Figure 2** Wind development and the PTC [11]

ten wind projects. Table 5 shows the amounts that were awarded to wind projects.

As shown in Table 6, European countries favor two mechanisms to facilitate renewable generation: the feed-in tariff (FIT) system and quota regulation in combination with a Tradable Green Certificates

(TGC) market. Other policy instruments, such as tender schemes, are no longer used as the main policy scheme in any European country. However, there are instruments, such as production tax incentives and investment incentives, which are frequently used as supplementary instruments [4]. FIT rates range from

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 5 ARRA wind project grants

Wind project awards	Location	Amount
Evergreen Wind Power V, LLC	Danforth, Maine	\$40,441,471
Moraine II Wind Farm	Woodstock, Minnesota	\$28,019,520
Canadaigua Power Partners, LLC	Cohocton, New York	\$52,352,334
Canadaigua Power Partners II, LLC	Cohocton, New York	\$22,296,494
Wheat Field Wind Farm	Arlington, Oregon	\$47,717,155
Hay Canyon Wind Farm	Moro, Oregon	\$47,092,555
Pebble Springs Wind Farm	Arlington, Oregon	\$46,543,219
Highland Wind Farm	Salix, Pennsylvania	\$42,204,562
Locust Ridge II, LLC	Shenandoah, Pennsylvania	\$59,162,064
Penascal Wind Farm	Sarita, Texas	\$114,071,646
Total awarded to wind plants		\$499,901,020

Source: DOE Press Release. *Treasury, Energy Announce \$500 Million in Awards for Clean Energy Projects*. September 1, 2009. <http://www.energy.gov/news2009/7851.htm>

€52 to €114/MWh for on shore wind plants and up to €130/MWh for off shore.

Green Power Markets

Green power markets can be classified as compliance and voluntary market. Compliance markets exist because of policy decisions, primarily state Renewable Portfolio Standards (RPS) that require electric service providers within the state to have a minimum amount of renewable energy in their electricity supply. RPSs also often define the specific types of renewable resources that are eligible for RPS compliance and include set-asides that direct suppliers to obtain a portion from a specific resource, most often solar power. For example, suppliers in New Jersey must meet a portion of their RPS from in-state solar generation; thus there is an internal

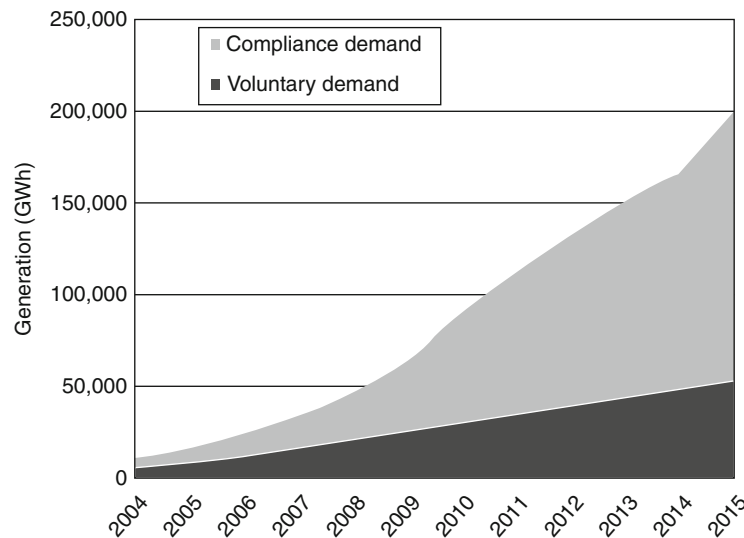
Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 6 Primary wind mechanisms in Europe

Primary mechanism for wind	European countries
Feed-in tariffs	Austria, Cyprus, Czech Republic, Estonia, France, Germany, Greece, Hungary, Ireland, Latvia, Lithuania, Luxembourg, Portugal, Slovakia, Slovenia, Spain, Switzerland
Quota obligation with TGC	Belgium, Italy, Poland, Romania, Sweden
Investment subsidies and tax incentives	Finland, Malta
Other	Bulgaria: mandatory purchase obligation
	Denmark: market price premium for on-shore and tendering system for off-shore
	The Netherlands – premium price tariff

New Jersey solar Renewable Energy Credit (REC) market for RPS purposes.

Voluntary markets are consumer driven and allow consumers to choose to do more than policy decisions require, that is, voluntarily purchase green power or RECs above and beyond their particular utilities RPS requirements. Voluntary consumer decisions to buy electricity supplied from renewable energy sources represent a powerful market support mechanism for renewable energy development. While the compliance market represents a larger portion of renewable energy and REC sales, the voluntary market helps develop nationwide renewable energy capacity that exceeds what mandatory markets contribute alone (see Fig. 3).

In the early 1990s, a small number of US utilities began offering “green power” options to their customers. Since then, these products have become more prevalent, both from traditional utilities and from renewable energy marketers operating in states that have introduced competition into their retail electricity markets or are offering RECs online. Today, more than half of all US electricity customers have an option to



Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Figure 3
Historic and projected demand for renewable energy [12]

voluntarily purchase some type of green power product directly from a retail electricity provider, while all consumers have the option to purchase RECs.

More than 850 utilities, or about 25% of utilities nationally in the United States, offer some type of green power programs to customers. These programs allow customers to purchase a green power product, which can include receiving some portion of their power supply as renewable energy (almost always at a higher price), offset a portion of their power supply with RECs, or contribute funds for the utility to invest in renewable energy development. The term “green power” generally refers to electricity supplied in whole or in part from renewable energy sources, such as wind and solar power, geothermal, hydropower (typically low-impact or small hydro), and various forms of biomass. Many utilities or alternative energy suppliers often do have access to local renewable energy resources but will offer a bundled green power product to consumers. A bundled green power product consists of the local electricity supply resource bundled with supplier purchased RECs.

In states with competitive (or restructured) retail electricity markets, electricity customers can buy electricity generated from renewable sources or bundled with RECs, by switching to an alternative electricity supplier that offers a green power product. In some of

these states, default utility electricity suppliers offer green power product options to their customers in conjunction with competitive green power marketers that do not require customers to switch. Nearly a dozen states that have opened their markets to retail competition have experienced some green power marketing activity.

Finally, regardless of whether they have access to a green power product from their retail power provider, any consumer can purchase green power by buying RECs in the voluntary markets, which represent the “environmental attributes” of electricity generated from renewable energy-based projects. This provides a way for consumers to support renewable energy development (through REC purchases) without having to switch to an alternative electricity supplier. Today, several dozen companies actively market RECs to residential or business customers throughout the United States. Many REC marketers also sell greenhouse gas emissions offsets sourced from renewable energy projects.

In 2008, approximately one million consumers and businesses purchased more than 24 billion kilowatt-hours (kWh) worth of green power products from their power suppliers or in the form of RECs (0.6% of total US 2008 electricity consumption). Of the total green power purchased, 71% was supplied from wind

energy, followed by biomass energy sources, including landfill gas (17%); hydropower (primarily low impact or small hydro) (9%); geothermal (2%); and solar (<1%). At the end of 2008, kilowatt-hour sales of renewable energy in voluntary markets represented a generating capacity equivalent of about 7,300 MW of renewable energy sources (1% of the peak US 2008 demand). The estimated market value of green power product sales in 2008 is between \$110 and \$190 million.

Green power product purchases have grown steadily in recent years, although preliminary data for 2009 suggest a slowing in the market presumably due to the economic downturn. In 2008, purchases increased by 34% (in kilowatt-hours) from the previous year, with annual average growth of 41% since 2004 (Table 7). REC purchases, primarily by businesses and institutional customers, have been driving much of the market growth, increasing 47% in 2008.

Purchases by nonresidential customers have outpaced those by residential consumers in recent years, with businesses and organizations now representing more than three-quarters of all green power product purchases. According to the US Environmental Protection Agency's Green Power Partnership, a program that works with businesses and organizations that purchase green power, some of the leading purchasers of green power, mainly in the form of RECs, in the United States include Intel, Kohl's, PepsiCo, Whole Foods, the City of Houston, and Dell. EPA maintains a list of the top purchasers on its Web site.

While full market data are not yet available for 2009, the EPA Green Power Partnership saw its partner purchases increase by 5% in 2009. This represents considerable slowing in partner purchases compared to the increases of 44% and 68% during the two previous years. Therefore, the economy is likely having some dampening effect on the market. In contrast, EPA also reported that it had about 300 companies or organizations join its Green Power Partnership in 2009, a large number based on historic standards, bringing the total Partners to 1,200.

Although renewable energy policies are now increasingly driving renewable energy development, voluntary markets can continue to play a role going forward. Today, more than half of US states have renewable energy standards requiring utilities or power suppliers to obtain a certain fraction of their generation from renewable sources (e.g., 20% renewable by 2020). However, consumers who want to purchase a greater share of renewable energy (100%) or support technologies that might not be incentivized by policies may want to continue to participate in these markets and support the development of new renewable energy resources.

Capacity Credit

Power system planners work to ensure that sufficient generating capacity is installed to meet projected demand for electricity. Although in some cases this

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 7 Estimated annual green power purchases by market sector, 2005–2008^a

Market sector	(Millions of kWh)							
	2005	2006	2007	2008	% Change 2004/2005	% Change 2005/2006	% Change 2006/2007	% Change 2007/2008
Utility green pricing	2,500	3,400	4,300	4,800	33%	39%	25%	12%
Competitive markets	2,200	1,700 ^b	3,200	3,900	–19%	–20% ^b	88% ^b	22%
REC markets ^c	3,900	6,800	10,600	15,600	126%	75%	55%	47%
<i>Retail total</i>	<i>8,500</i>	<i>11,900</i>	<i>18,100</i>	<i>24,300</i>	<i>37%</i>	<i>41%</i>	<i>53%</i>	<i>34%</i>

^aIncludes sales of new and existing renewable energy. Totals and growth rates may not calculate due to rounding.

^b2006 sales figures may be underestimated because of data gaps.

^cIncludes only RECs sold to end-use customers separate from electricity.

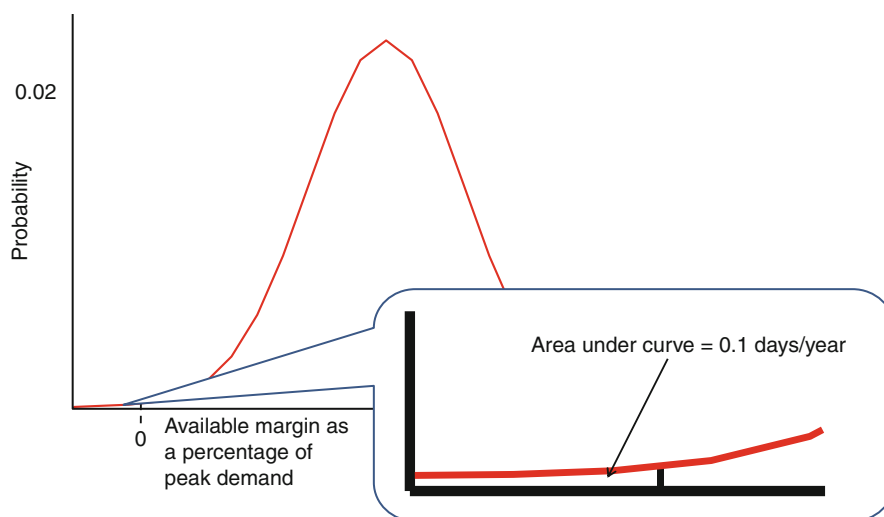
function is performed by markets, the fundamental issue of acquiring or otherwise procuring capacity is important to ensure long-term generation resource adequacy. To assess resource adequacy, a reliability-based analysis is usually performed that utilizes the forecast load demand, along with the generation fleets' reliability characteristics to determine whether there are sufficient generation resources at some future date. This typically uses probabilistic analysis so that the risk of having insufficient generation can be quantified. What is important is the overall system assessment, but it is often useful to assess the contribution that any individual power plant, or group of plants, makes toward overall resource adequacy. This contribution is known as the capacity credit, or capacity value, of the resource, and is typically measured either in MW of capacity or as a percentage of the rated capacity of the plant.

Calculating the capacity contribution of wind power plants requires detailed load data, data for the balance of generation, and wind production estimates, hourly or daily, that are time-synchronized with the load. The capacity value as measured by effective load-carrying capability is based on a reliability metric such as loss of load expectation (LOLE), which is often measured in units of days/year. A common reliability target is an LOLE of 0.1 day/year, or 1-day-in-10-years,

and is illustrated in the following figure as the area (magnified) under the left tail of the probability distribution in Fig. 4.

The basic principle underlying capacity value, called effective load-carrying capability (ELCC), is that the generation system should achieve the same reliability with wind as it does without wind. Using the above figure as an example, the generation fleet without the wind plant has an LOLE of 0.1 day/year. When the wind generation is added, the curve shifts to the right, indicating a high level of planning reserve during many, but not all, hours. This shift causes a reduction in the area under the curve, which was originally at 0.1 day/year, but is now less than that. To bring the system reliability back to its original value, the loads are increased incrementally until the area under the left tail of the curve is restored to its original value of 0.1 day/year. The amount of load increase that holds reliability constant is the ELCC of the wind plant.

The portfolio of generation needed to complement wind, especially at high penetration rates, is different than without wind. Typical capacity values for wind plants range from 5% to 40% of nameplate capacity, depending on the wind characteristics and correlation with load (values outside this range are possible, although somewhat unlikely). A conventional power plant can have an ELCC in the range of 50–95%,



Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Figure 4
Common reliability target: An LOLE of 0.1 day/year, or 1-day-in-10-years

whereas wind plants are significantly below this. Because of this, wind is primarily an energy resource that has limited capacity value. This implies that the generation portfolio that complements a high wind penetration scenario will need to provide a higher ratio of ELCC per unit of energy than would be the case without wind. For example, with a high wind penetration, instead of requiring a baseload capacity penetration of 50% of the peak demand (example only), the ratio of baseload generation would decline, and some baseload generation would be replaced by cycling generation with faster ramp rates (MW/min) and lower turndown levels (minimum sustained generation level, in MW).

Wind Integration Costs

Over the last several years, various entities have conducted wind integration studies in an effort to assess what is needed to accommodate larger amounts of wind energy on the electric grid, and how much this might cost. Integration studies generally simulate power systems under a certain penetration of wind energy and evaluate the impacts on the grid and estimate the incremental operating costs incurred due to the added variability and uncertainty of wind energy.

Two general approaches have been employed to determine wind integration impacts in the United States. The first approach estimates the cost of increased operating reserves in the regulation, load-following and unit commitment time frames that are required to balance the increased net load variability introduced by wind generation. These studies generally focus on wind's impact on ancillary services and compare wind with an energy-equivalent flat block of energy. The flat block will have no variability or uncertainty associated with it so that the operational differences can explicitly show the impacts of wind energy. Although the flat block is a useful construct because of its inherent certainty and constant energy delivery, recent work has identified some weaknesses in this approach [7]. Because the flat energy block is calculated to be energy-equivalent to the wind, large inter-day ramps in this artificial benchmark unit can occur as a result of significant differences in day-to-day production between the wind energy and the flat block. This effect can be significant at high wind penetrations.

In addition, the flat block schedule tends to have more on-peak energy and less off-peak energy than wind generation. As a result, the daily flat block is worth \$1.50–\$2.00/MWh more than the wind generation, and the resulting estimated wind integration costs may be overstated.

The second approach does not directly determine the cost of wind integration, but examines the impact that wind has on system commitment and dispatch, and calculates the net value of wind power between a base case and a case (or cases) with wind. The net value of wind represents cost savings from wind displacing more costly units, minus any additional costs that may be incurred from the need for additional ancillary services. Using this technique, the wind case is compared with that of an identical system without wind power, and all impacts are compared and considered from higher penetrations of wind power. Although this does not separate the impacts of wind on various levels of reserves, those impacts are captured implicitly in the modeling framework and results.

The technique most often employed is to estimate these costs as the difference in overall electric system production costs between a base case scenario and one that incorporates a certain level of variable and unpredictable wind energy penetration. Integration costs vary primarily due to the size of the balancing area being modeled, the accuracy of the wind forecasting (if used), the use (or not) of sub-hourly scheduling, the flexibility of the existing resource fleet, the type of ancillary services being employed, and the geographic diversity (or lack thereof) of the wind energy. The following table (Table 8) lists the main findings of the major US wind integration cost studies conducted between 2003 and 2007, most of which utilized some form of the flat block approach. The average cost of integrating wind energy was found to be about \$5.00/MWh. Recent work has raised the issue of whether existing methods for calculating integration costs are indeed measuring these costs accurately, and whether there may be integration costs for integrating other generating technologies that have not been addressed [8].

In 2008, the National Renewable Energy Laboratory (NREL), under the sponsorship of the U.S. Department of Energy, initiated the largest, most ambitious and comprehensive regional-level wind power integration studies conducted so far, in the United States.

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 8 Main results from major US wind integration cost studies

Date	Study	Wind capacity penetration	Integration cost (\$/MWh)				
			Regulation	Load following	Unit commitment	Gas supply	Total
2003	Xcel-UWIG	3.5%	0	0.41	1.44	–	1.85
2003	We energies	29%	1.02	0.15	1.75	–	2.92
2004	Xcel-MNDOC	15%	0.23	–	4.37	–	4.60
2005	PacifiCorp-2004	11%	0	1.48	3.16	–	4.64
2006	Calif. (multi-year) ^a	4%	0.45	Trace	Trace	–	0.45
2006	Xcel-PSCo	15%	0.20	–	3.32	1.45	4.97
2006	MN-MISO ^b	31%	–	–	–	–	4.41
2007	Puget sound energy	12%	–	–	–	–	6.94
2007	Arizona Pub. Service	15%	0.37	2.65	1.06	–	4.08
2007	Avista utilities	30%	1.43	4.40	3.00	–	8.84
2007	Idaho Power	20%	–	–	–	–	7.92
2007	PacifiCorp-2007	18%	–	1.10	4.00	–	5.10
2008	Xcel-PSCo ^c	20%	–	–	–	–	8.56
Average cost							5.02

Source: U.S. Department of Energy, 2008 *Wind technologies market report*, July 2009.

Note: not all types of costs were evaluated in all studies.

^aRegulation costs represent 3-year average.

^bHighest over 3-year evaluation period.

^cThis integration cost reflects a \$10/MMBtu natural gas price scenario. This cost is much higher than the integration cost calculated for Xcel-PSCo in 2006, in large measure due to the higher natural gas price: had the gas price from the 2006 study been used in the 2008 study, the integration cost would drop to \$5.13/MWh.

Western Wind and Solar Integration Study (WWSIS)

The Western Wind and Solar Integration Study (WWSIS) [9] examined up to 30% wind and 5% solar penetration scenarios in the WestConnect area, which includes Wyoming, Colorado, New Mexico, Arizona, and Nevada. This study evaluated system costs with and without wind energy, and also examined the implication of this high level of variable generation on reserves. Wind events occur much more slowly than contingency events. However, wind may increase the level of flexible reserves (ramping capability) that must be available, either as part of the unit commitment process or as quick-start capability that can be activated as needed. The International Energy Agency Task 25, Design of Power Systems with Large Amounts of Wind Energy, is examining this issue, as is the National Renewable Energy Laboratory in the United States in

collaboration with University College, Dublin's Electricity Research Center. This study did not attempt to estimate the integration cost per se of wind or solar, in part given the difficulty of properly establishing a robust benchmark case (Milligan and Kirby Integration Cost vs. value paper). However, the study did examine the economic impact of balancing area cooperation (joint unit commitment and economic dispatch) and found that the study footprint would experience savings of about \$2 billion/year (based on \$42 billion/year production cost) by utilizing this form of cooperation.

Eastern Wind Integration and Transmission Study (EWITS)

The Eastern Wind Integration and Transmission Study (EWITS) study [10] area includes most of the

Eastern Interconnection and is examining a 20% wind penetration scenario along with a transmission expansion component. The study includes the Midwest ISO, PJM, ISO New England, the New York ISO, the Southwest Power Pool, the Tennessee Valley Authority, the Mid-Continent Area Power Pool, and various other interested parties.

The EWITS study pushed the state of the art for wind integration cost forward. This study found that severe inter-day ramps, imposed by the flat-block proxy resource (which is only a fictitious benchmark unit) significantly distorted production costs. Instead, the EWITS approach estimated the cost based on adding wind into the system with no additional reserves as the “base case,” then determining the delta value by adding reserves to cover the additional variability and uncertainty of wind at the various penetration rates and geographic dispersion. In 2009 dollars, the integration costs ranged from about \$3.80 to \$5.40/MWh of wind generation.

Cost Trajectory for Wind Energy Technology

Historically, the busbar cost of wind energy has declined significantly, though it has increased in recent years. Although there are a number of factors behind this reduction in cost, the primary driver has been the adoption of new technologies that allow for lighter components and taller towers. This results in the capture of more energetic winds that tend to be higher from the earth’s surface. Today’s typical wind turbine has a tower height of 80–100 m. The historical cost decline has been driven not only by higher towers, but also by larger rotors that increase energy capture. However, it is likely that wind turbines will not continue to grow indefinitely because of the so-called square-cube law. This relationship says that as rotors increase in size, the energy capture will increase as the square of the swept area. At the same time, the volume of material required increases as the cube of the rotor size. Thus, cost increases faster than energy capture, holding constant all other factors. The square-cube law has been overcome by the introduction of advanced materials and other strategies that can reduce the per unit weight, but delivering the required volume, of material. As wind turbines become larger, there are also increases in logistical issues such as transporting large blade or

tower components across the land. For these reasons, it seems unlikely that the historical cost decline can continue, and that more modest incremental improvements will occur instead.

The logistical limitations of on-shore wind technology, transportation, are not as much of an issue with off-shore technology because ships can easily transport larger blade and tower units. However, the installation and operation/maintenance cost of off-shore turbines are higher because of the harsher sea-based environment, and the difficulty of accessing turbines for maintenance. Cost trajectories are uncertain, and will be driven in large part by the reliability of off-shore turbines. This issue, critical for all wind turbines, is especially significant because harsh weather conditions can limit access to off-shore turbines; therefore to mitigate the need for on-site maintenance a reliable turbine is necessary.

Table 9 is taken from Thresher et al., which is the source of much of this discussion, and shows results from the U.S. Department of Energy’s WindPACT Project. The table shows the likely ranges of cost and energy capture for future wind technology. Using 2002 as a base, the most likely outcome is a substantial increase in capacity factor with a small drop in capital cost, although the latter appears to be in doubt given the recent surges in commodity prices.

Wind Energy and Economic Development

Installing new wind power generation not only creates electricity, but also supports new manufacturing, construction, maintenance, and other jobs, as well as contributing to the economy. Depending on the local workforce and resources, a new wind project can provide many construction jobs for the local area, increase sales for local manufacturers and retailers, and support jobs and increased economic activity at local hotels, restaurants, grocery stores, and other suppliers of goods and services.

How much the new wind projects will impact their local community depends on many factors:

1. Are the developers using local labor for siting, legal assistance, road building, and turbine erection?
2. Will any of the project supplies or turbine components come from the local area (e.g., local blade production)?

Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Table 9 Areas of potential technology improvement

Areas of potential technology improvement			
Technical area	Potential advances	Cost increments (best/expected/least, percent)	
		Annual energy production	Turbine capital cost
Advanced tower concepts	* Taller towers in difficult locations	+11/+11/+11	+8/+12/+20
	* New materials and/or processes		
	* Advanced structures/foundations		
	* Self-erecting, initial or for service		
Advanced (enlarged) rotors	* Advanced materials	+35/+25/+10	−6/−3/+3
	* Improved structural-aero design		
	* Active controls		
	* Passive controls		
	* Higher tip speed/lower acoustics		
Reduced energy losses and improved availability	* Reduced blade soiling losses	+7/+5/0	0/0/0
	* Damage tolerant sensors		
	* Robust control systems		
	* Prognostic maintenance		
Drivetrain (gearboxes and generators and power electronics)	* Fewer gear stages or direct drive	+8/+4/0	−11/+6/+1
	* Medium/low speed generators		
	* Distributed gearbox topologies		
	* Permanent-magnet generators		
	* Medium voltage equipment		
	* Advanced gear tooth profiles		
	* New circuit topologies		
	* New semiconductor devices		
	* New materials (GaAs, SiC)		
Manufacturing and learning curve	* Sustained, incremental design and process improvements	0/0/0	−27/−13/−3
	* Large-scale manufacturing		
	* Reduced design loads		
Totals		+61/+45/+21	−36/−10/+21

- When the project is operational, will the owners hire local workers to maintain and repair turbines?
- What type of payments will landowners receive for allowing the turbines on their land?

- How are property taxes determined for the wind project owners?

Some of these questions are specific to the United States where typically, wind power developers will lease

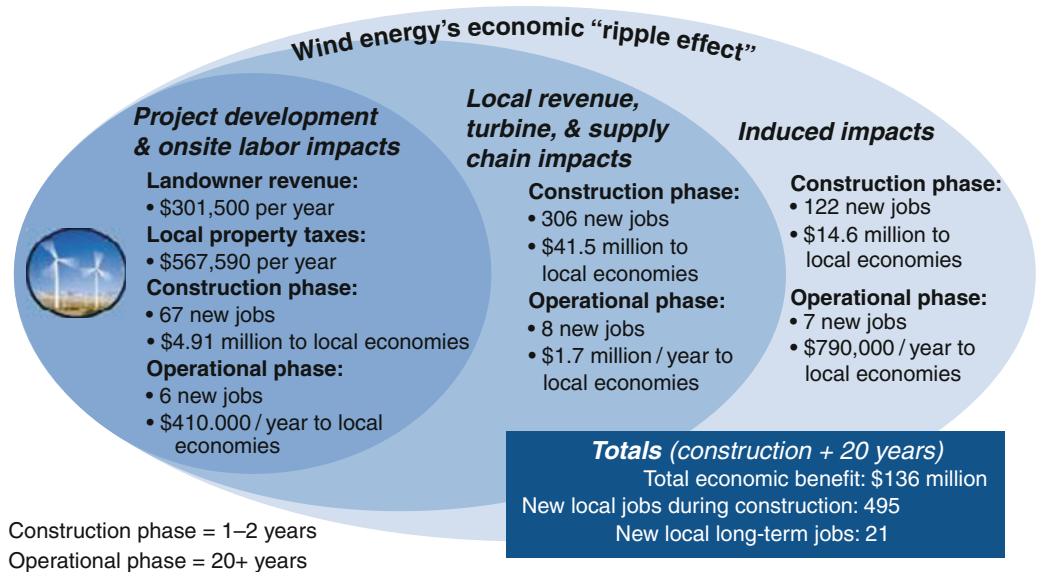
land from private citizens, and pay taxes to local counties, or negotiate with local governments for payments in lieu of taxes. However, the basic principles of economic development impacts apply everywhere. The net impact that wind energy has on jobs can be calculated by tracing the spending patterns and jobs created that are caused by the spending on the wind plant, and deducting the spending and jobs that would have been developed in the absence of wind. Approaches that can be applied include econometric analysis and input–output analysis, both of which are well-established methods. Input–output analysis can easily be adapted to alternative industries or locations and can be based on economic coefficients that describe the interaction between related sectors of the economy.

The JEDI Model

The U.S. Department of Energy’s National Renewable Energy Laboratory (NREL) has an input–output model that estimates the jobs and economic impacts associated with new wind power. The Jobs and Economic Development Impacts (JEDI) model requires very little user input, and is used by local officials, wind power

developers, and other interested parties to estimate local jobs, salaries, and economic development impacts associated with new wind projects. JEDI is based on project-specific inputs from the model user (the more information about the project, the more accurate the results will be) and on default inputs derived from industry averages and interviews. Below is a sample listing of jobs included in the JEDI output (Fig. 5).

JEDI is a free tool that provides the user with default information on land lease and property tax revenues for wind. Various ownership and financing structures can be incorporated by the user if they do not want to use the defaults. JEDI presents gross project-specific results. It does not consider potential electricity price impact or alternative investment options. JEDI results are reported on a statewide scale; however, JEDI can be used on a county, regional, or national basis by incorporating additional data (not included in the model). JEDI model defaults are based on interviews with industry experts and project developers. Economic multipliers contained within the model are US-specific and derived from a company called the Minnesota IMPLAN Group who uses data from many sources including the US Census and Bureau of Economic Analysis. You can learn more at www.implan.com



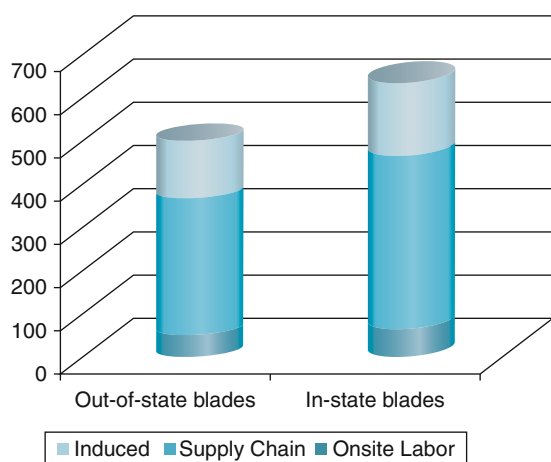
Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Figure 5
Wind energy's economic ripple effect

In addition to the JEDI wind model, there are also models for other technologies such as concentrating solar power and natural gas. These US-based models can be found at <http://www.nrel.gov/analysis/jedi/>

What Is the Difference with Local Components?

Typical US wind projects do not have locally made major components (blades, towers, and gear boxes). However, if any of the major components can be manufactured in the local region, the benefit to the local economy is much greater. This table shows the results of jobs and economic impacts from two almost identical projects. However, the first one has no locally made components and the second one has locally manufactured blades. This example is for the state of Colorado, where wind turbine blade manufacturing facilities are located.

Figure 6 displays results from the JEDI model, and shows that there will be approximately 140 more local jobs supported by the same new wind project if locally manufactured blades are used. Not only will the number of jobs increase but the economic impact to the (usually rural) community where the wind farm is placed will experience temporary economic growth during construction and permanent increases in operations jobs, land lease payments, property tax revenues, and local service jobs.



Wind Power: Economy, Market, Subsidies, Payment Mechanisms, and Capacity Credit. Figure 6
Jobs supported by new wind project: local vs. out-of-state blades

Future Directions

As the penetration of wind energy increases on power systems around the world, there remain questions regarding the interaction of wind energy with wholesale electricity markets. Wind energy has a near-zero marginal cost, which provides it with a high degree of economic merit in economic dispatch decisions that rank generation by marginal cost. At high penetration rates, wind may significantly reduce the spot price for electricity, which will reduce payments to conventional generation owners. In more extreme cases, the revenue shortfall for other generators may make it impossible for owners to meet capital cost payments. This may discourage base-load generation that cannot be easily or quickly cycled offline, instead encouraging more flexible generators that can economically operate at lower capacity factors. However, it is not clear, nor has it been conclusively demonstrated, that an economically efficient outcome can be attained in the long run. It is possible that capacity, along with other ancillary services, may be required in the short run to ensure reliable and economic operation of the power system; and in the long run to elicit the required level of flexibility in the balance of the power system. Power systems with high levels of wind generation will generally need more flexible generation (increased ramping and lower turn-down capability), access to this flexibility through market or other institutional mechanisms, and operational practice that can make best possible use of the technologies and institutions involved.

Various types of subsidies or requirements for minimum levels of renewable energy are attempts to incorporate all costs (private costs plus external costs) into decision making and implementation. Because pollution cost falls on a third party, these attempts at market adjustment, if done correctly, can elicit the economically efficient level of renewable energy. However, experience in the United States shows that tax credits (and other nonmarket incentives) can play havoc with development if wind energy if the credits longevity are in question.

Future areas of research will likely cover several broad themes: technology development, analytic tools and metrics for integration, and electricity market development.

Land-based turbine developments include advanced tower concepts, advanced rotors, reduced energy losses,

drive-train/gear-box reliability improvements, and increased efficiency and lower costs in manufacturing. Each of these evolutionary directions is expected to lower the bus-bar cost of wind energy, thus improving the economics. Off-shore turbine technology developments will also include more advances in platform designs and costs, improving the economics of this variation in wind energy technology.

There is increasing interest in wind plants that can provide reserves to the power system. Much of the required technology exists, although institutional constraints and lack of general awareness of this capability has hindered its use. With increasing wind penetrations, better institutional/operating rules, and modest changes in control algorithms, wind turbines could soon provide much of the reserves that are now required from conventional generation to help support wind energy.

At high wind penetrations there are questions about whether the existing markets will induce the required levels of flexibility from the remainder of the generation fleet. Research and analysis, including advanced power system market modeling will help answer these questions, and suggest potential future enhancements to markets that may include more robust capacity markets, metrics to assess flexibility need and supply, and relevant market products over the key time scales most relevant to wind energy integration. At the same time, advanced methods for unit commitment that involved stochastic optimization, although promising, will not be easily integrated into power markets without significant development.

Bibliography

Primary Literature

1. Milligan M, Kirby B, Gramlich R, Goggin M (2009) Impact of electric industry structure on high wind penetration potential. NREL/TP-550-46273. National Renewable Energy Laboratory, Golden. <http://www.nrel.gov/docs/fy09osti/46273.pdf>
2. ISO/RTO Council (2007) Increasing renewable resources, how ISOs and RTOs are helping meet this public policy objective. ISO/RTO Council, 16 Oct 2007
3. ENTSO (2009) The harmonized electricity market role model Version 2009–01. <http://www.entsoe.eu/>. Accessed 26 July 2009
4. Kjaer C et al (2009) Wind energy – the facts. European Wind Energy Association, Brussels
5. Morthorst P-E, Awerbuch S (2009) The economics of wind energy. European Wind Energy Association, Brussels
6. Morthorst PE (April 2007) TradeWind: further developing Europe's power market for large scale integration of wind power. Risø National Laboratory, Technical University of Denmark
7. Milligan M, Kirby B (2009) Calculating wind integration costs: separating wind energy value from integration cost impacts. NREL Report No. TP-550-46275. National Renewable Energy Laboratory, Golden, p 28. www.nrel.gov/publications
8. Milligan M, Ela E, Lew D, Corbus D (2010) Advancing wind integration study methodologies: implications of higher levels of wind. NREL technical report (to appear)
9. <http://www.nrel.gov/wind/systemsintegration/wwwis.html>
10. <http://www.nrel.gov/wind/systemsintegration/ewits.html>
11. Wiser R, Bolinger M (2008) Wind technologies market report. NREL Report No. TP-6A2-46026. DOE/GO-102009-2868. Department of Energy, Energy Efficiency & Renewable Energy Office. Golden, July 2009
12. Bird et al (June 2010) An examination of the regional supply and demand balance for renewable electricity in the United States through 2015. National Renewable Energy Laboratory, Golden
13. Wiser R, Bolinger M, Barbose G (November 2007) Using the federal production tax credit to build a durable market for wind power in the United States. Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley

Books and Reviews

- Bird L, Kreycik C, Friedman B (2009) Green power marketing in the United States: a status report (2008 Data). NREL/TP-6A2-46581. National Renewable Energy Laboratory, Golden <http://www.nrel.gov/docs/fy09osti/46581.pdf>. Accessed Sep 2009
- U.S. DOE Green Power Network Web site: <http://apps3.eere.energy.gov/greenpower/> U.S. EPA Green Power Partnership <http://www.epa.gov/greenpower/>

Wind Turbine Noise Emissions

KARL BOLIN, MATS ÅBOM
KTH Royal Institute of Technology, KTH-The Marcus Wallenberg Laboratory for Sound and Vibration Research, Stockholm, Sweden

Article Outline (Table of Contents)

Glossary
Definition of the Subject and Its Importance
Introduction

Basic Concepts
 Noise Sources
 Sound Propagation
 Human Impact
 Conclusions
 Future Directions
 Bibliography

Glossary

Sound pressure (p) The amplitude of the fluid's pressure disturbances around the steady state pressure which in air is perceived as sound, see Fig. 1.

Noise Is unwanted sound.

Sound pressure level (SPL, L_p) A logarithmic function of the sound pressure with the unit decibel [dB], see (7)

Sound level (L_A , L_B , L_C) Filtered sound pressure level with different filters (A, B, C) resembling the human hearing system at different loudness, see Fig. 6.

Acoustic power level (L_W) A measure of the radiated sound power in dB from acoustic sources, see (8).

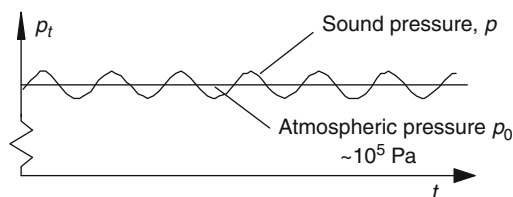
Sound speed (c) The propagation speed of sound in different media.

Audible sound Defined as pressure fluctuations between 20 Hz and 20 kHz.

Amplitude modulation Sound with a periodically varying level (amplitude).

Masking The change of loudness of one sound due to another.

Low-frequency sound Is normally defined as audible sound between 20 Hz and 100 Hz or 200 Hz.



Wind Turbine Noise Emissions. Figure 1

An acoustic pressure field $p(\mathbf{r}, t)$ ordinarily implies only small disturbances. Normal speech at a distance of several meters, for example, gives a sound pressure p of the order 0.01 Pa superimposed upon atmospheric pressure p_0 , which is about 10^5 Pa

Directivity The variation of sound pressure level in different directions from a source

Definition of the Subject and Its Importance

Wind turbines generate sound. Mainly this sound source is of aeroacoustic origin as improved sound insulation of the nacelles has reduced the mechanical sound sources. Unwanted sound is called noise and is a major factor when assessing nuisance caused by wind turbines. The increasing height of wind turbines has led to better possibilities to harvest wind at higher altitudes but has also contributed to increased noise levels in stable atmospheric conditions when the sound is easily transmitted due to a high wind speed gradient. Rapid development of large wind farms in rural, earlier unexploited, areas could cause large acceptance problems for nearby residents if noise is not properly considered in the development process. Several epidemiological studies have shown noise is the most prominent annoyance and show strong interdependence to visibility and lack of economic incentive from wind power.

Introduction

Wind turbines symbolize the hopes of a brighter future and solutions to serious problems such as global warming and energy shortage. Today's rapid expansion of wind turbines, both in quantity and size, can have a dramatic impact on the landscape in rural areas throughout the world. Although greenhouse gas emissions of wind turbines are low, the local scenery is often altered when wind farms are constructed. This often leads to concerned neighbors forming pressure groups opposing the projects. Transparent and unbiased information from contractors can often reduce the negative reactions from neighbors. The nearby residents are also usually protected by noise guidelines from authorities that developers have to comply with.

Naturally, not all wind energy is transformed into electricity at wind turbines. Inevitably some power is lost, for instance, in the generator or due to friction-induced forces in the mechanical system. Some of these losses generate sound. Early reports on wind turbine noise originate from the 1980s [1, 2]. Today's wind turbines mainly emit sound from the blades as they move through the air [3]. The amount of sound transmitted through the atmosphere is governed by not only

the acoustic reflection at the ground but also by the atmospheric conditions [4, 5]. Sound from wind turbines regularly causes annoyance to people nearby and should therefore be perceived as noise or unwanted sound. Noise from wind turbines is the main cause of concern to nearby residents [6–8] and should therefore be considered when determining the environmental impact of wind turbines.

Basic Concepts

Sound can be seen as an elastic wave motion where pressure disturbances propagate in a gas or a fluid with a sound speed c . At a certain point in space $\mathbf{r}$ and time t , an acoustic wave can be described by the sound pressure $p(\mathbf{r}, t)$ [Pa]. Sound pressure variations are normally very small deviations of the ambient pressure as illustrated in Fig. 1.

Omitting the effects of humidity and atmospheric pressure, the speed of sound in still air is given by

$$c = 331 \sqrt{\frac{273 + T}{273}} \quad [\text{m/s}], \quad (1)$$

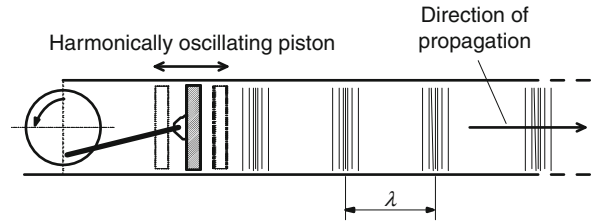
where T is the temperature in degrees Celsius. If there is wind present, the propagation is affected and the effective sound speed becomes: $c_{\text{eff}} = c + U$, where U is the flow speed in the direction of the sound propagation. For outdoor sound propagation, the temperature and wind speed variations in the atmosphere will cause sound speed variations. These variations will have an effect on long distance sound propagation since it will cause the sound waves to refract and propagate along curved paths, see section “Sound Propagation.”

The *root mean square* or *RMS value* is marked by $\tilde{p}$ and is defined according to

$$\tilde{p} = \sqrt{\frac{1}{T} \int_0^T p^2(t) dt}, \quad (2)$$

where T is the integration time. The rms-value is very important since it gives information about the time average of the power in the acoustic wave.

Plane waves are characterized by the condition that points with the same acoustical state form parallel planes. Figure 2 shows how a pure plane wave is excited in a duct by a harmonically oscillating piston at one end.



Wind Turbine Noise Emissions. Figure 2

A harmonically oscillating piston in an infinitely long duct gives rise to a plane longitudinal acoustic wave that propagates in the duct. The frequency f of the wave is the same as the piston oscillating frequency. The wavelength λ , i.e., the distance between two pressure maxima or minima, is related to f and the sound speed via: $c = f\lambda$.

For outdoor sound propagation, plane waves are useful, since at large distance from a source (*far field*), all sound fields can be approximated as plane waves. In order to characterize the strength of an acoustic source, such as wind turbines or fans, its (time-averaged) acoustic power W is measured for a certain operating condition. For a plane wave, the (time-averaged) acoustic power per unit area or intensity I is given by

$$I = \tilde{p}^2 / \rho_0 c, \quad (3)$$

where ρ_0 is the density and c the speed of sound (for air at 20 °C and normal pressure $\rho_0 c$ is around 400 [Pa s/m]).

Spherical and Cylindrical Sound Waves

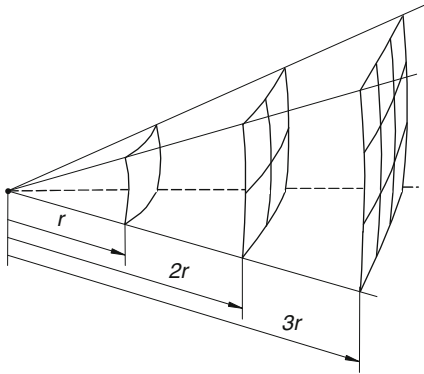
When modeling outdoor sound sources, two simple models are spherical and cylindrical waves. A spherical model is useful at large distances from any source and implies that the (time-averaged) sound power W is spreading equally over a large spherical surface, see Fig. 3.

This implies that intensity for a spherical source is given by

$$I = \frac{W}{4\pi r^2}, \quad (4)$$

where r is the radial distance. In this equation, the intensity for a plane wave can be used at large distances from the source which leads to

$$\tilde{p}^2 = \frac{\rho_0 c W}{4\pi r^2}. \quad (5)$$



Wind Turbine Noise Emissions. Figure 3

In spherical wave propagation, sound power is divided over an ever-increasing area. The intensity decreases to one fourth of its original value for a doubling of the distance to the source, and to one ninth when the distance is tripled

This last equation can be used to compute the sound pressure from a source with a given power at a certain distance. Alternatively, it can be used to determine the sound power by measuring the sound pressure and this is actually the method used for wind turbines. In practice, an outdoor source is normally on a reflecting ground, and to compensate for this, propagation over a half sphere is assumed. This means that the factor 4π is changed to 2π in (4) and (5). Another modification of (5) is to allow that the sound or acoustic intensity emitted in different directions varies. This variation of the sound intensity is called *source directivity*.

Similarly by assuming propagation over a cylindrical surface with height (or width) H , one can derive the formula

$$\tilde{p}^2 = \frac{\rho_0 c W}{2\pi r H}. \quad (6)$$

This equation applies at large distances for outdoor sound sources which can be seen as a line source (a road) or at some distance for a source located between two reflecting surfaces. An example would be a sea-based wind turbine, where the sea surface represents a totally reflecting surface and where, for certain atmospheric conditions, sound propagating upward is also reflected below a certain maximum height H [9]. When used for large distance, the effect of dissipation

during the propagation must also be considered in (5) and (6), see section “[Attenuation and Geometrical Spreading](#).” This will introduce an exponential factor $[\exp(-\alpha r)]$ in these equations, where the frequency-dependent damping constant α can be obtained from ISO 9613.

Levels and Decibels

At the beginning of the 1920s, it had become possible to carry out routine sound measurements. An acoustic group at Bell Systems in the USA introduced a measurement quantity that they called a “sensation unit,” which was based on a logarithmic scale with base 10. The unit *Bel* (after Alexander Graham Bell), defined as the base-10 logarithm ($\log_{10}$) of the quotient between two acoustic power values, eventually proved to be impractically large. Today, the unit commonly used is, instead, a tenth of a Bel. One decibel (1 dB) corresponds both to the measurement precision that can typically be obtained in acoustic measurements and to the amount of change that a human can discern in ideal circumstances.

The *Sound Pressure Level* L_p (or *SPL*) is defined as

$$L_p = 10 \cdot \log_{10} \frac{\tilde{p}^2}{p_{ref}^2} \quad (7)$$

where $\tilde{p}$ is the rms-amplitude of the sound pressure and $p_{ref} = 2 \cdot 10^{-5}$ Pa is the *reference value of sound pressure*. The reference value of sound pressure approximately corresponds to the lowest sound pressure that a young person with normal hearing can perceive at 1,000 Hz. As an example, ordinary speech corresponds to a sound pressure level of around 60 dB.

The *Sound Power Level* L_W indicates the acoustic power with respect to an internationally accepted reference as

$$L_W = 10 \cdot \log_{10} \frac{W}{W_{ref}} \quad (8)$$

where W is the time-averaged sound power and $W_{ref} = 10^{-12}$ [W] is the *reference value of sound power*.

Sound power is used to define the strength of various sources and from a known sound power and an assumption of the radiation, e.g., spherical waves, the sound pressure at a certain distance can be computed.

As an example for a source emitting spherical waves (5), one obtains the following relation between sound power and sound pressure levels

$$L_p = L_W - 10 \log_{10}(4\pi r^2) \approx L_W - 11 - 20 \cdot \log_{10} r, \quad (9)$$

where it is assumed that $\rho_0 c = 400$ [Pa s/m] and r is the distance in m.

Examples of sound powers and sound power levels from different objects can be seen in Fig. 4. From (9), it follows that at 1 m, the sound pressure level is 11 dB less than the sound power level and for each increase in distance with a factor 10, the level will decrease with 20 dB.

Addition of Sound Fields

For sound pressure, the principle of superposition applies; so, the total sound pressure $p_{tot}(t)$ is the sum of the individual waves sound pressures $p_n(t)$, i.e.,

$$p_{tot}(t) = \sum_{n=1}^N p_n(t). \quad (10)$$

In order to determine the SPL of the resulting total sound pressure, the RMS sound pressure must be used in accordance with (7). Assuming the most common

case that the individual waves originate from N independent sources, the total RMS value is simply

$$\tilde{p}_{tot}^2 = \sum_{n=1}^N \tilde{p}_n^2 \quad (11)$$

The corresponding rule for the total SPL from N sources becomes

$$L_{p_{tot}} = 10 \cdot \log_{10} \sum_{n=1}^N 10^{L_{pn}/10}. \quad (12)$$

As an example, if two sources of equal strength are added:

$$\begin{aligned} L_{p_{tot}} &= 10 \cdot \log_{10}(10^{L_{p1}/10} + 10^{L_{p1}/10}) \\ &= 10 \cdot \log_{10}(10^{L_{p1}/10} \cdot 2) \\ &= L_{p1} + 10 \cdot \log_{10} 2 = L_{p1} + 3 \text{ dB}, \end{aligned}$$

that is an increase of 3 dB in the SPL.

Characterization of Sound According to Frequency

Sound can be characterized by its frequency content. For mankind, *audible sound* is normally defined as the range 20–20,000 Hz. Frequencies lower than 20 Hz are called *infrasound*, and over 20,000 Hz are called *ultrasound*. Most acoustic phenomena are frequency-dependent,

Sound power level		Object	Acoustic power [W]
Acoustic power [W]	L_W [dB] ref 10^{-12} W		
100 000 000	200	Saturn rocket	50 000 000
1 000 000	180		
10 000	160	Four jetplanes	50 000
100	140		
1	120	Large orchestra Scream	10 1
0.01	100		
0.000 1	80	Typical speech	20 · 10 ⁻⁶
0.000 001	60		
0.000 000 01	40	Whispering	10 ⁻⁹
0.000 000 000 1	20		
0.000 000 000 001	0		

Wind Turbine Noise Emissions. Figure 4

The sound power and Sound Power Level for a number of typical sound sources

and it is well known that different sound sources have different characters, see Fig. 5.

Mathematically, each time signal, for instance, an acoustic pressure $p(t)$, can be seen as sum (integral) of harmonic signal. This is called Fourier analysis and means that an acoustic signal can be represented as a function of frequency or a spectrum. Normally spectra are expressed; so, they represent power which implies that for sound pressure, they are based on the RMS values. The total power in a spectrum can be obtained by applying (11) and sum over all the frequencies since different frequencies can be regarded as independent sources.

In practice, the frequency resolution in a Fourier analysis is related to the measurement time T , giving a resolution of $\Delta f = 1/T$. For noise control purposes and especially when studying sources with a broadband character, i.e., no distinct tones, it is convenient to present the data in broader frequency bands. Often one-third or octave bands are used for this purpose. The standardized octave bands covering the audible range are: 31.5, 63, 125, 250, 500, 1 k, 2 k, 4 k, 8 k,

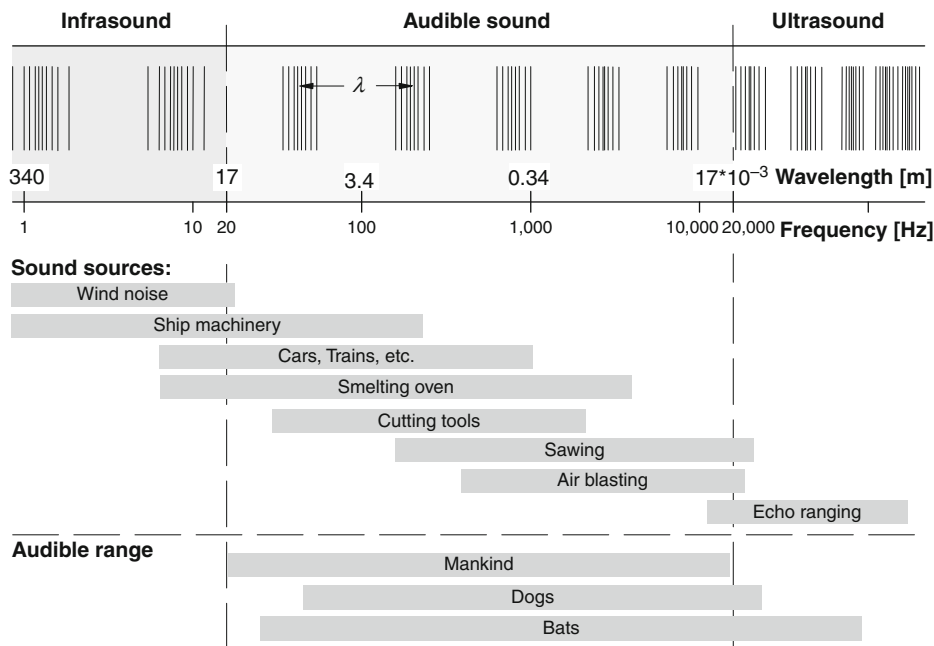
16 kHz. Each of these octave bands can be split into 3 one-third octaves.

Measures of Noise

The strength of sound is measured by a sound level meter that, in its simplest form, gives the SPL in dB. The SPL does not, however, take account of the nonlinearity of our perception with respect to frequency, as reflected in the concept of loudness [10]. To better reflect the human perception of sound, sound level meters contain filters, so-called *weighting filters*, that amplify the microphone signal in different amounts at different frequencies; see Fig. 6.

Today A-weighting is the common measure for characterizing outdoor sound and most regulations refer to A-weighted dB values, sometimes denoted dB (A) or dBA. A sound pressure level which is A-weighted is called *sound level* and can be denoted L_A . If the spectrum of sound is known

$$L_A = 10 \cdot \log_{10} \sum_{n=1}^N 10^{(L_{pn} + A_{An})/10} \quad [\text{dB}] \quad (13)$$



Wind Turbine Noise Emissions. Figure 5

Classification of sound according to frequency, and the relation between frequency and wavelength in air with a sound speed $c=340$ m/s. Frequency ranges of various sound generation mechanisms

where L_{pn} [dB] is the third-octave or octave band level in band n and ΔA_n [dB] is A-weighting in band n .

Determination of Wind Turbine Source Data

To characterize a wind turbine, the acoustic power or sound power level L_w (8) is estimated using a standard procedure, which in most countries comply with the IEC 61400-11 [11]. This procedure stipulates that the measurements are done at a reference position on the ground at different wind speeds. The wind speed data used should preferably be taken at the center of the turbine (the hub). The measured data is then interpolated to correspond to a reference wind speed of 8 m/s at 10 m height, assuming a standard atmospheric wind profile. Normally manufacturers present the data for a wind turbine as an A-weighted sound power plus the octave band spectrum for the power. Given the A-weighted power, the sound level from a turbine up to a distance of 1,000 m can be estimated from [12]

$$L_A = L_{WA} - 8 - 20 \cdot \log_{10} r - 0.005 \cdot r \quad [\text{dB}], \quad (14)$$

where L_{WA} is the A-weighted sound power level, r the distance in m, and the last term represents atmospheric damping. Equation 14 is based on (9) modified for a half spherical wave, i.e., assuming a reflecting ground. A typical value for the sound power level for 1 MW machines is around 100 dB(A). It can be noted that in

terms of acoustic power output, this only corresponds to 0.01 W, i.e., a single wind turbine is not a powerful sound source. As a comparison, a large passenger aircraft at takeoff creates acoustic powers in the range 1,000–10,000 W.

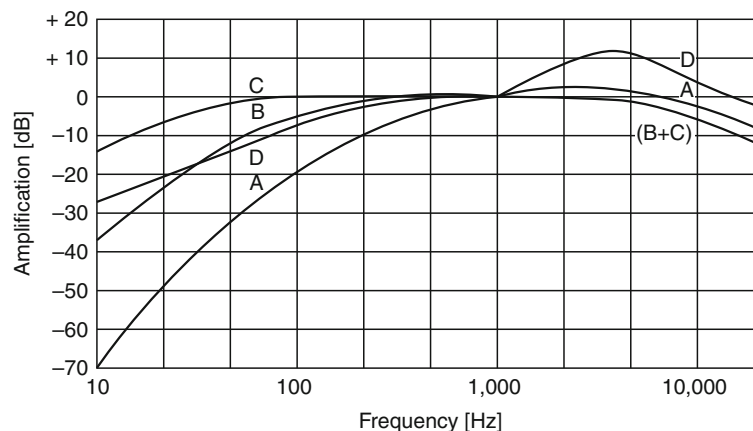
Noise Sources

Most modern wind turbines are of the type shown in Fig. 7, i.e., consist of a rotor attached to a horizontal axis which via a gearbox drives a generator. The rotor + gearbox and generator are enclosed in a housing which is mounted on a tower. Other less common types are based on a vertical rotor and axis where the gearbox and generator can be put on the ground. Here the discussion will focus on the first most common type of turbines.

The noise sources on a wind turbine can be split into two types: mechanical and aerodynamic.

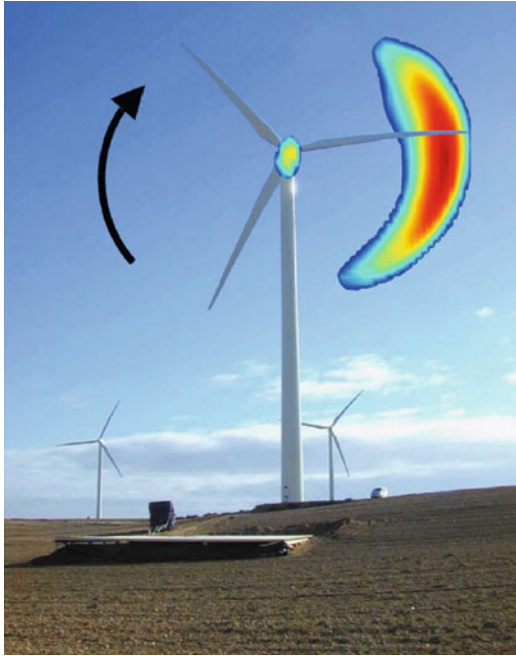
Mechanical Noise

The mechanical noise is normally dominated by the gearbox and generator, the strongest source being the gearbox, which transmits its vibrations to surfaces that can radiate sound to the surrounding, i.e., the housing, the rotor blades, and the tower. For modern well-designed machines, the mechanical noise can be considered to be of secondary importance compared to the



Wind Turbine Noise Emissions. Figure 6

A-, B-, C-, and D-weighting curves. A-weighting is the most common. For low and high frequencies, the amplification is negative, implying that these frequencies are damped to compensate for the lower sensitivity of humans to low- and high-frequency sound



Wind Turbine Noise Emissions. Figure 7

Example of modern 2.3 MW wind turbine with rotor diameter 94 m and hub height 100 m [13, 14].

In front (upwind) of the turbine is an acoustic antenna which produced the source image plot shown in the picture. The more “hot” a region is the stronger the acoustic source in the rotor plane

aerodynamic noise [13–15]. However, since gearbox noise can produce tones which are more annoying than the aerodynamic broadband noise, mechanical defects in a machine can lead to temporary noise problems.

Aerodynamic or Aeroacoustic Noise

Sound production by a flow can be explained by Lighthills equation [16]

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) p = \frac{\partial^2 \rho_0 u_i u_j}{\partial x_i \partial x_j}, \quad (15)$$

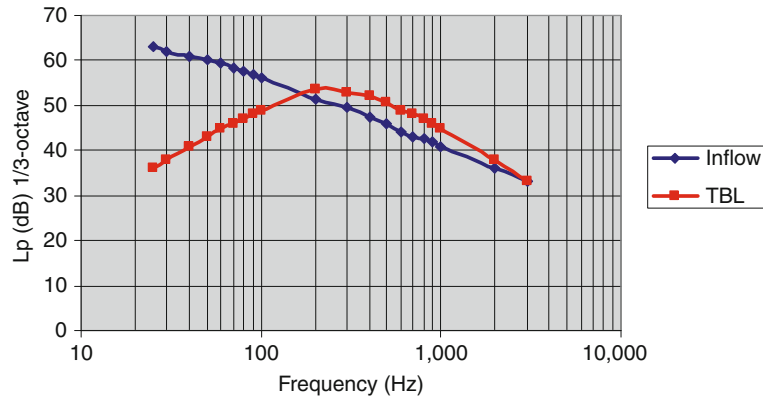
where u represents a flow velocity component, ∇^2 is the Laplace-operator, x is a Cartesian coordinate axis, $i, j = 1, 2, 3$ and a summation over i, j is assumed. The left hand side of (15) represents the classical wave equation and the right hand side is the flow-driven source distribution creating the sound field. The source term

represents an unsteady transport of momentum in the flow and is related to turbulence. When solid surfaces are present, the source term will be modified and can be amplified in particular at sharp edges [17]. Motion of the surface will also modify the resulting sound generation and cause Doppler (frequency) shift as well as change the directivity [16].

The aerodynamic noise can be split into three parts: low frequency, inflow-turbulence noise, and airfoil self-noise [13–15]. The low-frequency noise is related to interaction between the blades and tower wakes or large vortex structures from nearby rotors. The problem with tower wakes is uncommon nowadays when most rotors are operated upstream of the tower. The effect of vortex structures from nearby rotors can be a problem in wind turbine parks with too closely spaced machines. The inflow-turbulence noise is, for low frequencies, related to the modulation of the lifting force of the blade by the incoming turbulence which changes the flow speed and angle of attack. For higher frequencies, it can be seen as an interaction between the turbulence and the leading edge. The incoming turbulence will depend on the meteorological conditions as well as the surrounding terrain, i.e., one can expect quite different conditions for a turbine in open terrain as compared to one placed in a forest area.

The self-noise is the minimum sound a rotor will produce when placed in an ideal undisturbed inflow with minimum turbulence. The most important self-noise mechanism for large rotors is turbulent boundary layer trailing edge noise (TBL). Other possible sources are trailing edge bluntness noise and blade tip noise. But both these mechanisms are normally small for modern well-designed turbines and can be disregarded.

In conclusion, the turbulence-inflow and the TBL noise are the dominating aerodynamic sources. A typical example of the spectra produced by these mechanisms is shown in Fig. 8. The turbulence-inflow noise tends to be dominating at low frequencies less than 100–200 Hz for large wind turbines. The TBL noise has a distinct peak and dominates the midfrequency range which is important for the overall A-weighted sound pressure. The total (A-weighted) sound level is therefore normally dominated by TBL contribution but one cannot say that the turbulence-inflow part can be completely neglected [15]. An interesting note in this context is that for long range



Wind Turbine Noise Emissions. Figure 8

Example of computed inflow-turbulence and TBL noise spectra for a wind turbine with 61 m hub height and a rotor diameter of 91 m. The wind speed is 10.3 m/s at the hub and the rotor RPM 17.5. The model used is based on Lowson [18] where more details can be found

propagation (>5 km), the atmospheric damping will remove the mid- and frequency spectrum, which could make the inflow-noise contribution the dominating source.

Efforts to design silent blades and in particular to reduce the TBL noise have been made, for instance, in the EU-project DATA [19]. The result obtained by noise-optimized blade designs was an overall reduction of 4 dB. In addition, the idea [20] to modify a (straight) trailing edge, e.g., into use a saw tooth shape, to reduce the interaction between the turbulence and the edge was tested and gave an additional 2–3 dB reduction [19].

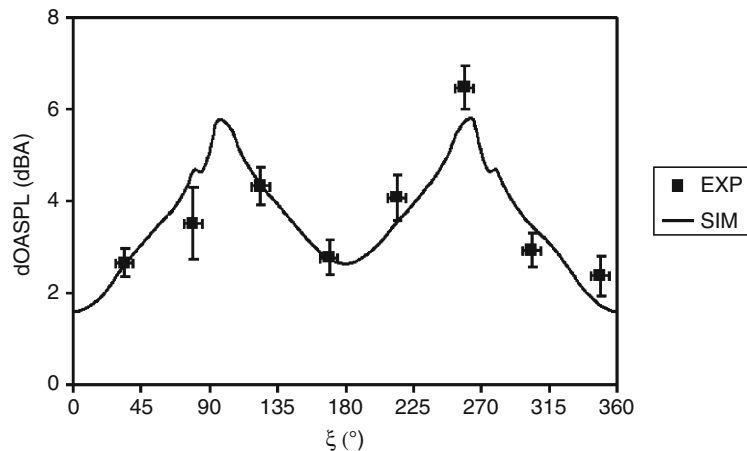
Source Directivity and Modulation Figure 7 shows a measured source distribution for a large wind turbine. It is clear from this picture that almost all sound radiated (to the ground) is emitted during the downward motion of the blades. The explanation for this unsymmetrical source distribution is associated with the directivity of the TBL noise and the effect of motion on this directivity [13, 14]. This directivity will cause a pulsating sound described by an amplitude modulation of the TBL noise with the blade passing frequency and is responsible for the characteristic “swish” sound close to wind turbines, see Fig. 9. As seen in the figure, the modulation amplitude just in front and behind a large wind turbine is around 2 dB(A). Positions in the rotor plane generate much larger modulations or close to 6 dB(A).

For a group of wind turbines which are running independently, modulation effects will tend to average out at larger distances. Also the fact the transmission through the atmosphere is affected by stochastic and time-varying conditions will tend to reduce the modulation.

Concerning the directivity in the ground plane, there is significant variation as shown in Fig. 10. As seen from the figure, there is an 8 dB(A) drop in level between the position right in front of the turbine and in the rotor plane.

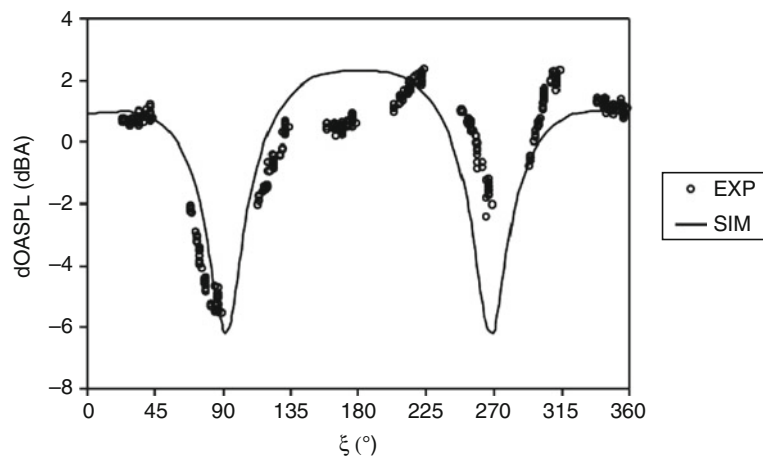
Sound Propagation

The perceived wind turbine noise is not only depending on the source but the transmission also plays an essential role. From a brief examination of the propagation problem, one might believe that sound is spherically dispersed and that the ground either absorb or reflect the sound rays. However, the nature of sound propagation is much more complex than this. Sound is a wave motion and consequently refracts when moving through regions with changing wave speed just like light when moving through a prism. Furthermore, ground reflections are altered both by the irregularities in the terrain as well as different amounts of reflection. Therefore, the sound propagation from wind turbines is varying with place and time. A turbine that is easily heard downwind of the plant in stable atmospheric conditions could be



Wind Turbine Noise Emissions. Figure 9

Far-field modulation amplitudes for TBL noise for the wind turbine in Fig. 7. The position 0° corresponds to right behind (downwind) and 180° right in front of the wind turbine (upwind). (From Ref. [14])



Wind Turbine Noise Emissions. Figure 10

Directivity in the far field for the wind turbine in Fig. 7. The position 0° corresponds to right behind (downwind) and 180° right in front of the wind turbine (upwind). (From Ref. [14])

impossible to perceive when the wind shift direction and the atmosphere changes to an unstable state.

Attenuation and Geometrical Spreading

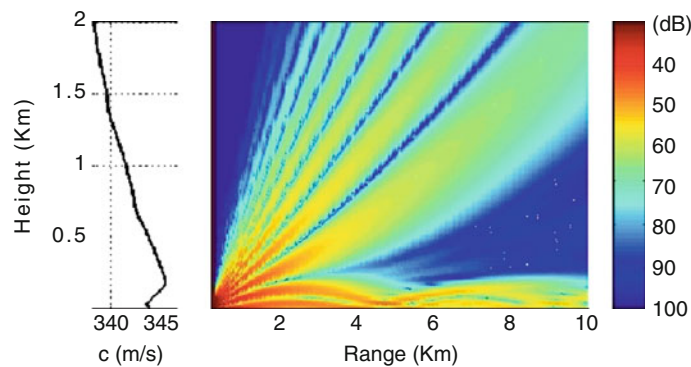
The intensity of the sound decrease as it propagates. This reduction is caused by geometrical spreading, see Sect. “Spherical and Cylindrical Sound Waves,” and by attenuation of the sound pressure.

Attenuation of sound is due to thermal and viscous effects as well as molecular relaxation when the sound wave propagates. The amount of attenuation is dependent on sound frequency, air temperature, atmospheric pressure, and humidity. The attenuation increases with frequency, atmospheric pressure, and humidity but decreases with temperature. A standardized calculation procedure, ISO 9613 [21], exists for frequencies from 20 Hz to 20 kHz.

Sounds are also geometrically dispersed as described in Sect. “Spherical and Cylindrical Sound Waves.” Close to the source, in the near field, this spreading is spherical and is proportional to r^2 . However, at longer distances, this rule of thumb could not be used as sound rays reflect at the ground and refract when traveling through air with different acoustic sound speeds. Examples of refracting sound waves can be seen in Figs. 11 and 12. As can be seen in the former figure, the sound is refracted downward and a channel of high sound intensity is observed below 200 m altitude. If instead the sound is refracted

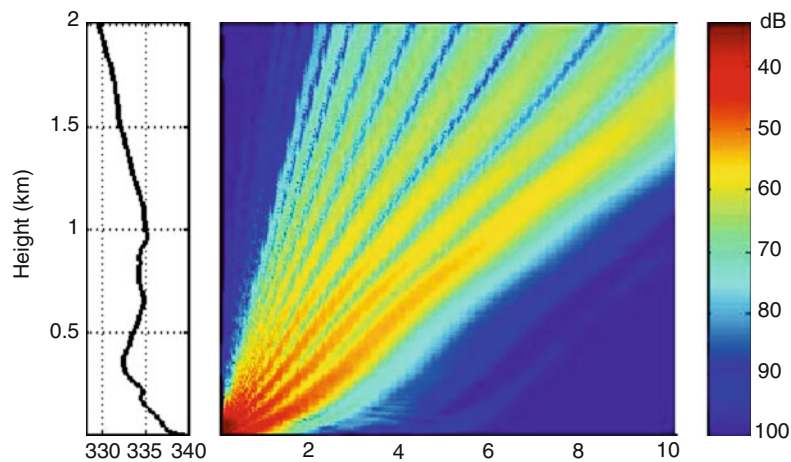
upward, as shown in Fig. 12, a sound shadow is observed on the ground after 2 km range. To further complicate the matter, wind turbulence scatters sound which causes the pronounced shadow zone in Fig. 12 to be diminished when taking the fluctuations of wind into account as can be seen in Fig. 13.

Ground reflections are also an important factor to take into account for sound propagation. In some conditions, the ground is absorbing sound, for example, when covered by snow or with leaves. On the other hand, at water surfaces or concrete ground, the sound is almost totally reflected. Measurements and models of



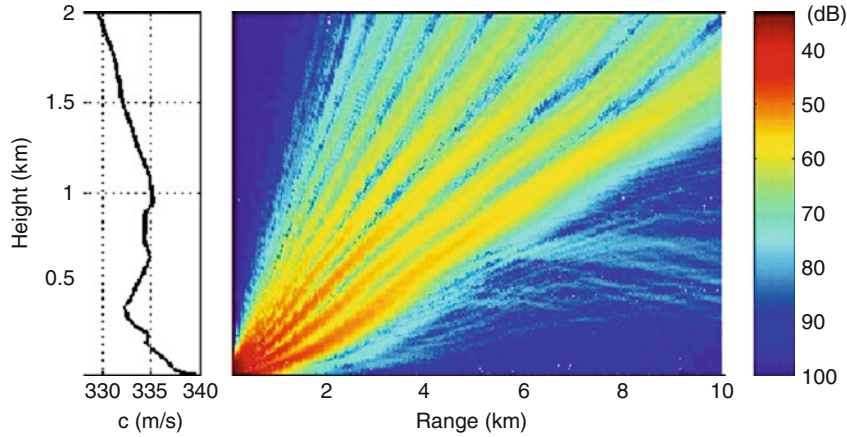
Wind Turbine Noise Emissions. Figure 11

Sound speed profile (*left*) and sound field (*right*). Computed sound transmission in a downward refracting atmosphere (increasing wind speed up to 200 m altitude) is shown. (Reprint from Bolin et al. [5])



Wind Turbine Noise Emissions. Figure 12

Sound speed profile (*left*) and sound field (*right*). Computed sound transmission in a laminar upward refracting atmosphere (decreasing wind speed) is shown. (Reprint from Bolin et al. [5])



Wind Turbine Noise Emissions. Figure 13

Sound speed profile (left) and sound field (right). Computed sound transmission in a turbulent upward refracting atmosphere (decreasing wind speed) is shown. (Reprint from Bolin et al. [5])

ground reflections from sound have been conducted by Delany and Bazley [22] and Attenborough [23] among others. Often, the terrain is irregular. In such conditions, the sound rays are not reflected with the same incidence angle as reflected angle in the horizontal plane as seen in Fig. 14. Periodical perturbations of the ground, for example sea waves, can under certain conditions be modeled according to Boss theory [24]. Furthermore, complex terrain such as hills and forests influences the atmospheric conditions [25] and consequently the sound propagation. For propagation through vegetation, an extra attenuation of the sound should be added as described in ISO9613 [21].

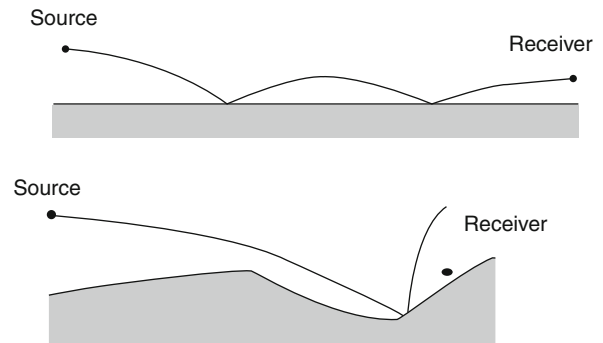
Sound can be seen as a longitudinal wave spreading from the source. The basic equation for the propagation of sound is consequently the wave equation without a source term (RHS of (16) equals zero) seen below

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial \tau^2} = 0 \quad (16)$$

The sound wave can be considered harmonic if the time dependence of the sound pressure can be described by

$$p(r, t) = \text{Re}(p_c e^{-i\omega t}) \quad (17)$$

where p_c is the amplitude, ωt is the phase where ω is the angular frequency ($f = 2\pi\omega$). Thus, the wave equation (16) can be rewritten into the Helmholtz equation



Wind Turbine Noise Emissions. Figure 14

Top: Sound ray in a downward refracting atmosphere reaching the receiver in two reflections. Bottom: Sound ray in irregular terrain not reaching the receiver

$$\nabla^2 p_c - k^2 p_c = 0 \quad (18)$$

where $k = \omega/c$ is the wave number. The field of computing sound propagation consequently largely deals with techniques to solve this equation.

Computational Atmospheric Acoustics

The field of computational atmospheric acoustics has many similarities to the fields of underwater acoustics as well as seismology. Although boundary conditions and input parameters differ depending on the field of

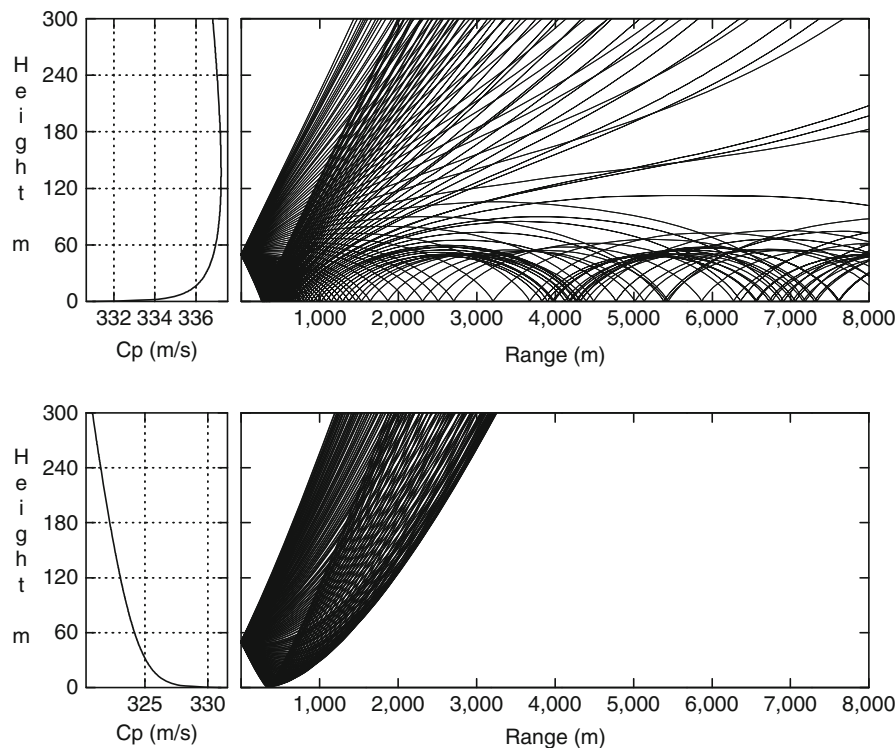
research, the governing equations are essentially the same. Therefore, numerical algorithms developed in one area are often also useful at other applications.

Methods of calculating atmospheric sound propagation can be divided into several subgroups. Here these different techniques are only covered briefly by comments on some of their advantages and disadvantages respectively. For an excellent overview of the field of atmospheric sound propagation, the interested reader is referred to Salomons [26]. Several different approaches exist to estimate sound propagation. Some of these methods are based on the Helmholtz equation (18), and is either direct solutions to this or using variable substitutions to solve the problem. On the other hand, other methods are based on ray tracing algorithms and Snell's law.

- *ISO9613* [21] is an engineering method to estimate sound propagation outdoors under “favorable”

sound transmission conditions. These situations are characterized by downwind and moderately downward refracting atmospheric states. Furthermore, the method is stated to be used for ground-based noise sources. However, wind turbines nowadays emit sound at much higher altitudes than the method was intended to cover. Although the algorithm can be considered coarse and somewhat inappropriate to be used for wind turbine applications, it is widely implemented in noise mapping, probably because of the simplicity and computational speed of the estimations or maybe because of the credibility due to being an ISO-standard.

- *Ray tracing* methods, also known as geometrical acoustics, are based on viewing the sound as refracting beams propagating from the source to the receiver. An example of a downward refracting sound speed profile can be seen in the upper subfigure of Fig. 15. The sound is in this case



Wind Turbine Noise Emissions. Figure 15

Upper: Sound speed profiles for increasing (upper figure) with height and decreasing (lower figure) with height. Lower: Pictures of the sound rays propagating from a source to 8 km distance

bent downward and the intensity is higher than for spherical spreading. This simulation can be compared to upward refracting conditions shown in the lower subfigure of Fig. 15. As can be seen, the sound energy is bent upward and an acoustic shadow zone occurs at the ground after 0.5 km. Apart from illustrative examples of ray tracing calculations, Fig. 15 also shows the two most important shortcomings of ray tracing methods. In the upper figure, points where sound rays converge can be observed. At these positions, ray tracing algorithms predict infinite sound pressures which is incorrect. In the lower figure, a shadow zone, a region without rays, is seen after 0.5 km at ground level. This prognosis is also erroneous and gives the false impression that no sound reaches this area while in fact sound propagates into this region. Methods, for instance, caustic diffraction fields [27, 28], deal with these two incorrect states but the ray tracing is still considered unsuitable for accurate sound propagation calculations at complicated sound speed profiles [26].

- *Fast field programs* were originally developed for underwater acoustic applications. This method is the numerical solution of the inhomogeneous Helmholtz equation for a monopole sound source

$$k^2 \frac{\partial}{\partial z} \left(k^{-2} \frac{\partial P}{\partial z} \right) + k^2 P = -4\pi\delta(z - z_s) \quad (19)$$

Where k is the wave number, P is the Fourier transform of p in the horizontal wave number domain, δ is the Dirac delta function, z_s is the source height. This algorithm can be used in both two and three dimensions but requires assumptions of constant ground conditions and wave numbers in (19) within each layer. As the reality often deviates strongly from the assumptions mentioned above, especially for long propagation distances, Fast field programs are suitable to deal with the sound propagation at short ranges. Analytical solutions of (19) are possible for very simple sound speed profiles and are referred to as normal mode solutions.

- *Parabolic equation solutions* are another family of useful computational methods for estimating sound propagation. If the Helmholtz equation (18) is assumed axisymmetric, it is reduced from

three to two dimensions. It can be shown [26] that by using the variable substitution $p = c^{-1/2}\psi^* \exp(ikr)$, the two-dimensional Helmholtz equation may be rewritten as

$$\frac{\partial^2 \psi}{\partial r^2} + 2ik(z) \frac{\partial \psi}{\partial r} + \frac{\partial^2 \psi}{\partial z^2} + (k(z) - k_0(z))\psi = 0 \quad (20)$$

neglecting $\partial^2 \psi / \partial r^2$ is possible if ψ varies slowly with range which is the case for narrow propagation angles which gives

$$2ik(z) \frac{\partial \psi}{\partial r} + \frac{\partial^2 \psi}{\partial z^2} + (k(z) - k_0(z))\psi = 0 \quad (21)$$

which has the shape of a parabolic equation. Several different numerical methods to solve this equation exist, for instance, Crank-Nicholson and Green's function solutions. These methods can use an arbitrary varying sound speed profile and changing ground conditions but are unsuitable for computations of the near field as the pressure field is not slowly varying in that region. However, at long ranges, such methods combined with local meteorological measurement have been shown to give accurate predictions of sound transmission at offshore wind turbines [5].

Concerning sound propagation calculations of wind turbine noise, these usually use the ISO 9613 [21] standard, and the acoustic source power for the turbine is acquired from the manufacturer at certain wind speeds according to the different national standards. This yields estimations of sound levels at reception points indifferent to ground irregularities and meteorological conditions. The current trend in assessing the sound propagation issue is to use more advanced prediction methods than ISO9613. One example is the Nord2000 software based on ray tracing algorithms which take into account different meteorological classes, and therefore, weather-dependent estimations of the sound levels at reception points can be calculated. This can be considered a natural development as databases with meteorological and ground conditions exist in many countries. Using these databases in softwares such as NORD2000, computing sound propagation gives improved quality of the noise mapping at wind turbine sites.

Human Impact

Sound from wind turbine is per definition not noise until someone hears it and is annoyed by it. Then the reactions differ from person to person. In today's society, noise pollution is a serious issue affecting a large part of the population. Noise can, apart from causing annoyance, also lead to other adverse effects such as sleep disturbance, tinnitus, and hearing loss [29]. Weak evidence also points toward an increased risk for myocardial infarction when exposed to traffic noise [30]. The noise levels for wind turbines are far too low to evoke tinnitus and hearing loss. Sleep disturbance caused by wind turbines has been investigated by Pedersen & Persson-Waye [6] where 16% of respondents reported interrupted sleep at exposure levels above 35 dBA.

Noise and the Auditory System

In the field of acoustics, noise is defined as unwanted sound. The perception of sound sources consequently determines if they are perceived as wanted or unwanted. It has been shown that persons with economic incentive in nearby wind farms are less annoyed by the sound than people not benefiting from the turbines [8]. The owner enjoys the gentle hum while his neighbor is distressed by the eternal swoosh and the interrupted scenery. Before the 1970s and the oil crisis, wind turbine farms were relatively small and thus noise seldom spread to neighbors. This might be one reason that noise annoyance was not a severe issue before the development of larger wind turbines in the 1980s when the adverse effects from noise were first observed by staff at NASA [1].

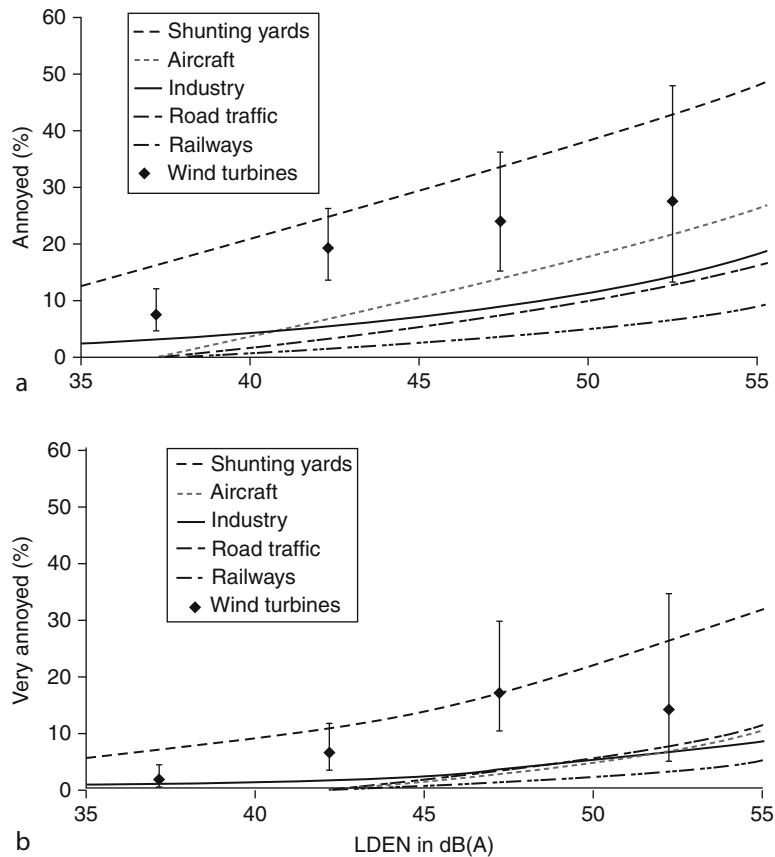
The auditory system in humans is a complicated and much investigated sensory organ. It starts at the ear lobe working as an antenna and then the sound propagates into the ear canal and reaches the eardrum. Here the pressure fluctuations are transmitted to the inner ear via the middle ear bones malleus, incus, and stapes, the smallest bones in the human body, to the inner ear and the cochlea where the organ of Corti transforms the movements in the fluid to nerve impulses. These are transmitted to the brain where they are processed and the sensation of hearing is produced. All this is indeed a remarkable system with an exceptional sensitivity,

recognizing pressure fluctuations down to 20 μ Pa and still robust enough to withstand amplitudes up to 400 Pa. It is partly due to this enormous flexibility the logarithmic dB scale is used in acoustics.

Dose–Response Relationships

It has been shown that increased wind turbine noise levels cause increasing annoyance to nearby residents. This correlation has been established in three different epidemiological studies by Pedersen and coauthors [6–8] performed in the Netherlands and in Sweden. Two studies have been conducted in flat farmland areas mainly of rural character [6, 8]. In the third paper [7] noise annoyance in both rural and suburban areas as well as hilly terrain was investigated. Apart from the statistically significant dose–response relationship, the studies also suggested that intrusive sound characteristics or visual impact influences the perceived annoyance [6]. It was also confirmed from an earlier study by Wolsink [31] that a negative opinion of the impact of wind turbines on the landscape has strong correlation to annoyance [8]. It should however be emphasized that it is not clear when negative opinions precede the annoyance or vice versa. Anyway, the results from the research performed by Pedersen and coauthors give extremely valuable insight when noise from wind turbines annoys nearby residents and should definitely be considered when developing wind turbine sites.

Comparisons between dose–response relationships of different community noise sources are seen in Fig. 16. As can be observed, wind turbine noise is considered more annoying than traffic noise at the same doses (dBA). This has many times been seen as a proof that wind turbine noise is much more annoying than most conventional noise sources. However, a problematic issue with these comparisons is the differing dose calculation between the sources. For most environmental noises, e.g., road traffic noise and aircraft noise, the noise dose for the affected people is estimated by traffic intensity and distance to source. However, for wind turbine noise, this is not the case. Rather than using the actual noise dose, the sound level at a specific wind speed (8 m/s at 10 m height) is used, thereby not knowing the actual noise dose of the subjects. This could give large differences in actual noise



Wind Turbine Noise Emissions. Figure 16

Comparison between dose-response relationship for other noise sources and wind turbine noise. Proportion of annoyed shown in (a) and highly annoyed in (b). (Reprint from [8])

exposure, for instance, old stalled regulated turbines increase their noise emission significantly from wind speeds at 8 m/s to maximum noise levels at cutoff wind speed as discussed in section “[Determination of Wind Turbine Source Data](#).” Consequently, at a site with high wind speeds, the dose for persons living close to such a turbine could be much higher than for others living at sites with lower wind speeds although both have the same nominal noise exposure at wind speed of 8 m/s at 10 m height.

Masking

When two sounds are heard simultaneously, the perception is altered as when heard alone. Therefore, the

loudness of a sound is not only dependent on its sound level but also on the context in which it is presented. This auditory phenomenon is called masking and will influence the perceived loudness of wind turbine noise.

For example, a turbine close to a busy highway will be less audible than the same turbine in a tranquil farmland.

Modern turbines usually have noise with broadband character without tonal components. As a result, ideal maskers are sources of similar character as spectral overlap between the sounds increases the masking effect. A laboratory study concerning if natural sounds could mask wind turbine noise has shown that inaudibility of wind turbine noise requires

10 dB higher background sounds than turbine noise [32]. Results from the same paper show that when the wind turbine and background have the same sound levels, the turbine is perceived as 5 dB quieter than if heard alone. In an epidemiological study, road traffic noise had to exceed the sound level of the turbines by 20 dB in order to show a statistically significant masking effect [8]. The British wind turbine noise guidelines [33] considers masking when setting emission levels by allowing a 5 dB higher equivalent sound level compared to the background noise exceeded 90% of the time. However, several factors influence the turbine noise and background sound. For example, traffic noise decreases in off-peak hours which often correspond to times when the turbine noise is loud. The increasing turbine heights give dramatic differences between the wind speed at hub height and at the ground [4]. This is especially pronounced in stable atmospheric conditions when the wind gradient increases. Thus, temporal variations, both of the wind turbine noise and the background sound, have to be considered if masking is to be evaluated correctly. For instance, wind-induced background sounds have to take fluctuations due to turbulence and different wind speed profiles into account [34]. Thus, although masking is probably commonly occurring, it remains to be evaluated in field conditions to what degree this has an effect in practice.

Low-Frequency Noise

When discussing low frequencies, acousticians define this as the lower range of the audible scale, often between 20 and 100 or 200 Hz. The growing size of wind turbines has generated fears that the low-frequency part has increased. Investigations of the impact at these frequencies suggest that the low-frequency levels have not increased significantly at newer turbines and it does not appear to be a source of concern [35, 36].

Guidelines

Different evaluation methods to regulate wind turbine noise are present all over the world. To give a full description of each of these is naturally impossible

within this book and therefore the principles governing these guidelines and some examples are treated in this section. The guidelines have been divided into three categories: the first based on absolute sound level limits, the second sort founded on relative sound level compared to background sounds, and the third kind with limits as a function of wind speed.

Sound level limits are the main method to base guidelines upon. This is the most obvious choice of limit and is practical and straightforward to comply with. Often these rules constrain the sound levels outside nearby residential buildings at specific meteorological conditions. For instance, the Danish guidelines are set to 45 dBA in residential areas at a wind speed of 8 m/s at 10 m height.

Limit values are often calculated in equivalent levels over a duration time T , denoted $L_{eq,T}$, which is an average of the sound pressure (normally A-weighted) expressed in dB. T is normally set to between 5 min and 1 h for wind turbine noise. It is often usual to have specific night time limits lower than day time values (see, for example, the German and State of Washington guidelines).

This sort of rule is simple and the relation between emission levels and noise annoyance has been established in epidemiological studies [6–8]. However, criticism of this kind of guidelines claims that the sensation of hearing is not only based on the level of the noise source but should be viewed in a wider context in order to get a more complete picture of the noise problem. Moreover, as discussed in section “[Dose–Response Relationships](#),” the noise dose at wind speeds of 8 m/s at 10 m height could deviate considerably from actual noise exposure.

Limits relative to background sound levels use the fact that masking of wind turbine noise by other sound sources influences the loudness. Background noise is usually wind generated, for instance, vegetation sounds [32]. Therefore, the allowed turbine noise depends on wind speed as in the British and French standards. Measurements of background sounds have to be performed at neighboring properties to apply this method. This kind of procedure can thus be time consuming as statistically reliable data should be gathered for different wind conditions and seasons.

Furthermore, a noise-polluted area, say from road traffic, will obtain more noise and this could be considered unfair. One way of circumventing both these objections is to use models for wind-generated noise, which exist for vegetation and sea wave sounds [34, 37]. However, the accuracy of these models is not yet verified by long measurement campaigns, and therefore, predictions can presently be seen as a complement to measurements rather than an independent base of information.

Levels as a function of wind speed can be considered equivalent to assuming uniform background levels at all locations. This is an intermediate between the two earlier described methods as sound level limits are set and masking assumed by increasing allowed limits with wind speed. From many points of view, this solution offers several positive aspects as masking is taken into account in a simple way. However, different locations have varying background sound and this is not taken into account. In this context, it is not surprising that the Netherlands has guidelines according to this principle as the country has a reasonably uniform terrain and therefore suitable to the required generalization.

Conclusions

Wind turbine noise has to be considered in the planning process when building wind farms. Sound is mainly described in the unit decibel (dB) which is a logarithmic function of the sound pressure. Noise is mainly generated by unsteady aerodynamic flows at the blades trailing edge, as improvements have reduced the mechanical noise sources and consequently tonal sounds are no longer a problem in current wind turbines (as of year 2011) when properly working. Today pitch-regulated wind turbines can also be run in silent modes where less noise generation is preferred rather than optimal power output. Sound transmission from source to receiver is dependent on the amount of ground reflections and also by the meteorological conditions. Acoustically hard ground, such as water, and atmospheric conditions of downward refracting sound increase the proportion of sound transmitted to receivers. Models dealing with sound propagation vary from relatively simple algorithms mainly valid at shorter distances to more complicated methods at longer distances and complex terrain. The wind turbine

noise effects on humans are mainly sleep disturbance and nuisance. Correlation between visibility and lack of economic incentive and annoyance has been reported from epidemiological studies. Different noise emission guidelines exist. Some of these specify a limit at certain wind speeds while others have limits depending on the background sound level. However, the context governs which approach and specific limits that are appropriate. To preserve the scenic beauty of landscapes and not add unnecessary noise to pristine environments, concern of the environmental impact in general and noise in particular is strongly advised when planning new wind farms. Noise emissions from wind turbines will persist, and with a large expansion of the number as well as size, the problem could increase drastically if proper measures are not taken in the planning stage. Therefore, noise should be considered at an early phase in the development process and an open dialogue between wind contractors and the public is strongly advised.

Future Directions

The rapid development of wind turbines will probably continue in the forthcoming years. The average size of wind turbines has increased tremendously over the last decades. If and when this growth will be halted is difficult to foresee. As not only the size but also the number of wind turbines is increasing, large populations, mainly in rural areas, will be affected by wind turbine noise. The potential for reducing the dominating aerodynamic noise by optimized blades is in the range 4–6 dB, which is not sufficient to solve the problem. Improvements of the noise mapping by advanced noise propagation algorithms are therefore important to minimize the noise emission in critical areas and to warn when noise annoyance issues are likely to arise. Furthermore, sleep disturbance tests would also be an interesting research subject as higher emission levels often occur at nighttime. Field studies regarding masking that would validate results from laboratory studies would also give interesting results. As a conclusion, the need for further studies regarding the noise problem is important as noise is the most common source of concern for residents near wind turbines.

Bibliography

- Stephens DG, Shepherd KP, Hubbard HH, Grosveld FW (1982) Guide to the evaluation of human exposure to noise from large wind turbines. NASA TM 83288, Langley Research Center, Hampton, Virginia, USA
- Sutherland LC, Mantey R, Brown R (1987) Environmental and cumulative impact of noise from major wind turbine generator developments in Alameda and Riverside Counties- Literature review and measurement procedure development, Wyle Research Report WR87-8
- Guidati G, Ostertag J, Wagner S (2000) Prediction and reduction of wind turbine noise: an overview of research activities in Europe, AIAA paper 2000-0042
- van den Berg GP (2004) Effects of the wind profile at night on wind turbine sound. *J Sound Vib* 277:955–970
- Bolin K, Boué M, Karasalo I (2009) Long range sound propagation over a sea surface. *J Acoust Soc Am* 126: 2191–2197
- Pedersen E, Persson-Waye K (2004) Perception and annoyance due to wind turbine noise. A dose-response relationship. *J Acoust Soc Am* 116:3460–3470
- Pedersen E, Persson-Waye K (2007) Wind turbine noise, annoyance and self-reported health and well-being in different living environments. *Occup Environ Med* 64(7): 480–486
- Pedersen E, van den Berg GP, Bakker R, Bouma J (2009) Response to noise from modern wind farms in the Netherlands. *J Acoust Soc Am* 126:634–643
- Åbom M, Boué M (2007) Long range sound propagation over the sea with application to wind turbine sound
- Zwicker E, Fastl H (2001) *Psychoacoustics: facts and models*. Springer, Berlin
- International Electrotechnical Commission (2001) IEC 61400-11 standard: wind turbine generator systems C part 11: acoustic noise measurement techniques
- Ljunggren S (2001) *Ljud från vindkraftverk (sound from wind turbines)*. Technical report 6241, Boverket, Energimyndigheten & Naturvårdsverket, Sweden (in Swedish)
- Oerlemans S, Sijtsma P, López BM (2007) Location and quantification of noise sources on a wind turbine. *J Sound Vib* 299:869–883
- Oerlemans S, Schepers JG (2009) Prediction of wind turbine noise and validation against experiment. Report National Aerospace Laboratory NLR-TP-2009-402
- Guidati G, Ostertag J, Wagner S (2000) Prediction and reduction of wind turbine noise: an overview of research activities in Europe. American Institute of Aeronautics and Astronautics Paper 200-0042
- Howe M (2008) *Acoustics of fluid-structure interactions*. Cambridge University Press, Cambridge
- Williams JEF, Hall LH (1970) Aerodynamic sound generation by turbulent flow in the vicinity of a scattering half plane. *J Fluid Mechanics* 40:657–670
- Lowson MV (1994) Design prediction model for wind turbine noise. Flow Solutions Report 94/06, ETSU W/13/00317/REP
- Oerlemans S, Schepers JG, Guidati G, Wagner S (2001) Experimental demonstration of wind turbine noise reduction through optimized airfoil shape and trailing-edge serrations. Report National Aerospace Laboratory NLR-TP-2001-324
- Hove MS (1991) Noise produced by a sawtooth trailing edge. *J Acoust Soc Am* 90:482–487
- ISO9613 (1995) Attenuation of sound during propagation outdoors. International Organization for Standardization, Geneva, Switzerland
- Delany ME, Bazley EN (1970) Acoustical properties of fibrous absorbent materials. *Appl Acoust* 3:105–116
- Attenborough K (1985) Acoustical impedance models for outdoor sound surfaces. *J Sound Vib* 99:521–544
- Boulanger P, Attenborough K, Taherzadeh S, Waters-Fuller T, Li KM (1998) Ground effect over hard rough surfaces. *J Acoust Soc Am* 104:1474–1482
- Kaimal JC, Finnigan JJ (1994) *Atmospheric boundary layer flows*. Oxford University Press, New York
- Salomons E (2002) *Computational atmospheric acoustics*. Kluwer, Dordrecht
- Ludwig D (1966) Uniform asymptotic expansion at a caustic. *Commun Pure Appl Math* 19:215–250
- Kratsov YA (1968) Two new wave asymptotic methods in the theory of wave propagation in inhomogeneous media. *Sov Phys Acoust* 14:1–17
- Berglund B, Lindvall T, Schwela DH (2000) *Guidelines for community noise*. World Health Organisation, Geneva
- Selander J, Nilsson ME, Bluhm G, Rosenlund M, Lindqvist M, Nise G, Pershagen G (2009) Long-term exposure to road traffic noise and myocardial infarction. *Epidemiology* 20: 272–279
- Wolsink M, Sprengers M (1993) Wind turbine noise: a new environmental threat? vol 2. ICBEN, Nice, France, pp 235–238
- Bolin K, Nilsson ME, Khan S (2010) The potential of natural sounds to mask wind turbine noise. *Acta Acustica united Acustica* 96:131–137
- Meir R et al. (1996) The assessment and rating of noise from wind farms. Technical report ETSU-R-97, Department of Trade and Industry, Harwell, UK
- Bolin K (2009) Prediction method for wind-induced vegetation noise. *Acta Acustica united Acustica* 75:607–619
- Pedersen T, Madsen KD (2010) Low frequency noise from large wind turbines, summary and conclusions on measurements and methods, Technical report Delta acoustics, AV1272/10
- O'Neal R, Hellweg RD, Lampeter RM (2011) Low frequency noise and infrasound from wind turbines. *Noise Control Eng J* 59(2):135–157
- Bolin K, Åbom M (2010) Air-borne sound generated by sea waves. *J Acoust Soc Am* 127:2771–2779